From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

---

Doctoral Dissertations, 2011-2026 Psychology

2020

Commonsense Aesthetics

Roek, Aaron Kurosu

Princeton, NJ : Princeton University

---

http://arks.princeton.edu/ark:/88435/dsp01kd17cw90r

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

COMMONSENSE AESTHETICS

Aaron Kurosu Roek

A DISSERTATION

PRESENTED TO THE FACULTY

OF PRINCETON UNIVERSITY

IN CANDIDACY FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE BY

THE DEPARTMENT OF PSYCHOLOGY

Adviser: Alexander Todorov, Ph.D.

September 2020

© Copyright by Aaron Kurosu, 2020. All rights reserved.

Abstract

Aesthetics is an old enigmatic topic in psychology. Here I focus on the beauty of every impression with an emphasis on the visual domain. My approach incorporates new tools (e.g., procedural generation, shape-families, webGL, virtual reality, 3D printing) and multivariate statistics (e.g., mixed-effects regression, PCA) to discover new insights. The purpose of my dissertation is to encourage further research elucidating the influence of shapes in the formation of shared impressions.

The first project focuses on aesthetics in terms of preferences–specifically, people's tendency to prefer curvier and smoother shapes. This aesthetic bias is one of the oldest phenomena studied in psychology. My contribution to the existing literature links the aesthetic bias to an objective definition for curvier and smoother shapes. The results suggest that people are sensitive to subtle objective differences in shapes that of which can affect preferences.

The latter two projects focus on aesthetics beyond preferences and more broadly in terms of any impression (e.g., artificial, dangerous, beautiful). I show that impressions of novel objects constitute reliable phenomena that could be further understood by studying shapes in more ways than traditional explored (e.g., curviness). I test stimuli consisting of sets of novel objects maintaining mathematically defined relationships called shape-families because object within them are similar and yet different. Using

shape-families, I describe the reliability of object impressions and explore the range of those that may be universally shared.

This dissertation represents a new look at aesthetics and shapes. For example, the prevailing focus of aesthetics today is on preferences or the experience of art. However, this work builds towards a focus on aesthetics in terms of agreement and focuses on a definition of aesthetics that may be generalizable to everyday experiences. Also, when compared to the dominant focus in vision research on studying shapes as it relates to object recognition, this work focuses on impressions. The cumulative results presented in this dissertation demonstrate evidence that there is meaning among the shapes of novel objects.

Acknowledgements

It takes a lot to get to this point. I am not just referring to the work I had put into writing this dissertation, or my focusing strategies to finish it despite the ongoing pandemic and protests. It takes a support group made up of family, friends, colleagues, and institutions. I want to acknowledge the support group that allowed me to focus these past six years on researching aesthetics.

Thank you to my committee members: Adele Goldberg, Uri Hasson, Casey Lew-Williams, & Diana Tamir. Thank you, especially to my advisor, Alex Todorov, for taking a chance on an unusual student with a design background. Thank you, Princeton University, for providing me the opportunity to study aesthetics earnestly. Thank you to my undergraduate institution, the University of Cincinnati, for introducing me to design and psychology. Thank you to my high school teacher, Liane Voss, for supporting my curiosity.

Thank you to my colleagues and friends. From before college: Christine Perez, Nikki Katsutani, Janelle Hiu, Justin Hedani. From my undergraduate years: Alex Broerman, Allison Dunphy, Anouk Moulliet, Ben Hsieh, Carlie Barrett, Casey Schneider, Chris Solada, Drew Alderfer, Jacklyn Ady, Jon Panichella, Matt Gill, Mei Hsieh, Michael Ruffing, Mike Roller, Natalie Sulpizio, Pooja Kadambi, Stephen Speaker, & Taylor Beeghly. During my graduate program: Brandon Labbree, Brandy Briones, DongWon

Oh, Dominic Burkart, Gandalf Nicolas, Joel Martinez, John Wolfe, Judith Mildner, Judy Fan, Mai Nguyen, Nina Rouhani, Stefan Uddenberg, Yeon Soon Shin, & Zidong Zhao.

Thank you to my family for supporting me every step of the way; Bob Tanaka, Craig Kurosu, Edie Watanabe, Evan Kurosu, Kathlyn Higa, Kathy Tanaka, Jan Narikawa, Masao Kurosu, Owen Narikawa, Roger Watanabe, & Yuni Kurosu. Extra special thanks to my husband, Greg Henselman-Petrusek Roek, and my biggest fan, my grandmother, Gladys Kurosu.

Table of Contents

Title page i

Copyright ii

Abstract iii

Table of contents vii

1. Introduction 1

2. Aesthetic shape continuity. 4

3. Reliability of object impressions. 29

4. Shared impressions of shapes. 76

5. Conclusion 138

References 142

Appendix 166

1.

Introduction

What about aesthetics is common sense? In Kantian terms (e.g., Hayn-Leichsenring & Chatterjee, 2019), I argue that aesthetics may be *universal* and exist as a *judgement of taste*. The final project included in this dissertation demonstrates this arcane notion empirically. To be honest, it was never my intention to test aesthetic philosophy. My research has been driven by pragmatic concerns. For example, the last project was actually motivated by a desire to understand how other people share impressions of novel objects. As someone who went to design school, I could never understand why some of my colleagues made such ugly things and called them beautiful.

There are three projects in this dissertation. The first project focuses on one of the oldest topics studied in psychology— curves. The *curvature effect*, as it is known today, is a general preference people have for curves over angles (Palumbo & Bertamini, 2016). It follows a classic approach to studying aesthetics: one that quests for universal principals; i.e., objective properties of beauty (e.g., Berlyne, 1970; Graham & Field, 2008; Swami & Furnham, 2012). What stood unknown in this old literature was how to

define curviness objectively to reproduce the effect. My contribution was serendipitous because I was actually motivated to test an analogous design principle perpetuated by industrial designers. It was a principle that, to my knowledge, had not been empirically tested, nor had it been linked to the research on the curvature effect. Borrowing from industrial design, the answer was geometric continuity. Using this mathematical definition to define curviness, the project demonstrates a replication of the curvature effect and adds something new to one of the oldest topics of research in psychology.

Before introducing the subsequent two projects, it should be mentioned that the study of aesthetics takes many forms in psychology. This is perhaps because aesthetics has been a subject of interest for a long time. For example, the German psychologist Fechner (1801-1887) ran the first psychophysical experiments while examining the aesthetics of rectangular proportions (i.e., Golden ratio) (M Bertamini & Palumbo, 2015). Prior to the advent of experimental psychology, philosophers such as Socrates were deeply pondering the essence of beauty (Hergenhahn, 2008). The aesthetic discourse today is diverse. Yet since the beginning of the 20th century, the dominant focus of aesthetics has been on art (Kranjec & Skov, 2020). For example, some researchers focus on understanding the experience of art while presupposing that the beauty of a painting and the beauty of an everyday object are disparate (Marković, 2010).

To be clear, this dissertation does not focus on the aesthetic experience of art. The first project focuses on aesthetics in terms of preferences. The latter two projects focus on aesthetics in terms of impressions. Focusing on aesthetics in terms of impressions represents a niche perspective more often explored by industrial designers (e.g., Mastandrea et al., 1992; Nagamachi, 1995).

Again, in the second chapter, I share a project examining aesthetics in terms of universal preferences for greater curvature or surface continuity. Then in the following third and fourth chapter, I share two projects examining aesthetics in terms of impressions formed from viewing shapes.

In chapter three, I first explore how an impression can be elicitable by the shape of novel objects. For many experimental psychologists, the primary implication is that different novel objects cannot be treated as neutral in meaning. For industrial designers, the project lends empirical support for their presumed ability to transform a shape to engender a specific impression. This may seem strange to a non-industrial designer, but just imagine—why it is that a face you have never seen before can look more or less trustworthy (e.g., Oosterhof & Todorov, 2008)? The basic idea is the same—designers believe that they could make an object look more or less trustworthy to someone who has never seen it before (or knows what they are looking at). The third project subsequently explores the types of impressions people share of novel objects. In Kantian terms, aesthetics in terms of impressions is similar to the notion of *free beauty*, because it is the experience of non-representational and non-art stimuli (e.g., Hayn-Leichsenring & Chatterjee, 2019).

In summary, the three projects in this dissertation span two different definitions of aesthetics. The first project examines aesthetics in the classic sense—it is about exploring a universal principal of beauty. The latter two projects examine aesthetics in a sense that focuses on impressions, i.e., feelings and thoughts in response to visual stimuli.

2.

Aesthetic shape continuity.[1]

2.1 Introduction

The curvature effect is an aesthetic bias to prefer curvier and smoother shapes (Marco Bertamini et al., 2016). There are centuries of observations on this topic (Gómez-Puerto et al., 2018). In 1753, for example, the British painter Hogarth theorized that some curves are more appealing than others. In fact, evidence for the curvature effect is robust and has been replicated in different contexts such as car interiors (Leder & Carbon, 2005), furniture (Dazkir & Read, 2012), packaging (Salgado-Montejo et al., 2015; Westerman et al., 2012), portraits (Hevner, 1935), and architecture (Banaei et al., 2019; Ruta et al., 2019; Shemesh et al., 2017; van Oel & van den Berkhof, 2013; Vartanian et al., 2013, 2019). The curvature effect does not appear to be tied to any specific category of objects. Instead, the effect appears to be a response to a quality of an object's shape.

[1] Work based on an unpublished manuscript by Kurosu, A., & Todorov, A.

Despite its long history and widespread applicability in industrial design, a psychophysical definition of the aesthetic bias remains unknown. Prior research on the curvature effect has primarily focused on *why* this bias in preference exists. For example, some researchers have suggested that the preference for curvature is really an outgrowth of an avoidance response towards threatening stimuli (e.g., Bar & Neta, 2006). However, recent research contends that threat alone is not a sufficient explanation (Marco Bertamini et al., 2016). Other researchers have explored the potential biological origins of the curvature effect. Those results demonstrate the bias existing among infants (Fantz & Miranda, 1975; Jadva et al., 2010; Quinn et al., 1997), great apes (Munar et al., 2015), and across cultures (Gómez-Puerto et al., 2018). Taken together, the curvature affect appears to be a robust experience-independent behavioral response.

What remains unknown in this literature is an objective definition of curviness. Researchers have, for example, relied on stimuli organized into *a priori* or normative categories to test the aesthetics of curviness (Bar & Neta, 2006; Marco Bertamini & Sinico, 2019; Guthrie & Wiener, 1966). A significant drawback to this approach is that some researchers have used semantically related, yet different words to form their categories (e.g., round v. sharp, curvy v. angular) (Gómez-Puerto et al., 2016). While there can be agreement in how words map to shape-parameters, this agreement does not always occur (A. Kurosu & Todorov, 2017). An alternative approach has been to have participants draw beautiful or ugly shapes and then categorize their curviness *post hoc* (Lundholm, 1921). In either approach, however, the reliance on a human definition, rather than an objective definition, is a significant limitation to this literature. An

objective definition would help to disentangle the contributions of semantics and visual information to the curvature effect.

We use a mathematical definition of continuity to operationalize curvature and identify the specific features that give rise to the curvature bias. In computer aided geometric design, continuity refers to the cross boundary derivative of two Bézier (i.e., polynomial) curves or surfaces (G. E. Farin & Farin, 2002). If you imagine running a finger along a line, continuity defines the transition of your finger from one position to the next. Industrial designers typically categorize continuity into three types: positional (POS), tangential (TAN), and curvature (CUR) (Wronecki, 2007); see **Figure 1**. These categories refer to the potential number of derivatives at the boundary between two curves or surfaces. POS refers to zero derivatives and defines angular shapes. TAN refers to one derivative and defines curvy and/or smooth shapes. CUR refers to two or more derivatives and defines smoother shapes relative to TAN. For more information, refer to the method section of this paper. From lowest to highest derivatives, adopting the notion of continuity provides a systematic nomenclature for understanding the curvature effect.

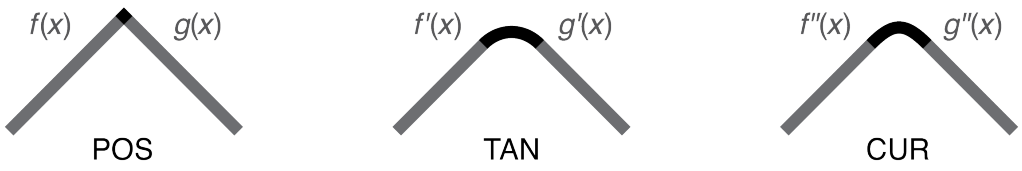


Figure 1. A diagram of continuity presented as two conjoined Bézier curves, $f(x)$ and $g(x)$. The transition by which they meet is the primary manipulation of this study.

We hypothesize that 'continuity' is the objective definition researchers have been searching for. It solves the tricky puzzle of finding a suitable definition that can replicate the curvature effect across any mode of visual presentation. The healthy girth of prior research demonstrates that the curvature effect is mode-independent; replicating with hand-drawn lines (Corradi et al., 2019; Hevner, 1935; Lundholm, 1921), 2D geometry (Marco Bertamini et al., 2016; Palumbo & Bertamini, 2016; Silvia & Barona, 2009), images of 3D objects and scenes (Corradi et al., 2019; Leder & Carbon, 2005; Munar et al., 2015; Silvia & Barona, 2009; van Oel & van den Berkhof, 2013), stereoscopically presented virtual architecture (Banaei et al., 2019; Shemesh et al., 2017), and real geometric solids (Soranzo et al., 2018). We predict that continuity can define shapes that replicate the curvature effect regardless of how they are presented visually.



Figure 2. Example of three types of continuity. The transition between the top surface and the conical surface has been manipulated to feature POS, TAN, or CUR. Together this represents one shape-family.

To test the aesthetics of shape continuity, we created a set of stimuli controlled to feature POS, TAN, or CUR; for an example, see **Figure 2**. We hypothesized that people would generally prefer greater continuity. Specifically, we predicted that TAN and CUR are generally preferred over POS. We also predicted CUR is generally preferred over TAN if people are sensitive to the subtle differences between these two types of continuity. To test the robustness of preferences for smoother objects, we tested five

different modes of presentation. The five modes presented the shapes as lines, images, animations of 3D objects on a 2D screen (mono-animate), animations of 3D objects in virtual reality (stereo-animate), and real objects, see **Figure 3**.



Figure 3. Five studies examining different presentation modes: contours (lines), static renderings (images), animations (mono-animate), virtual reality (stereo-animate), 3D-prints (real).

2.2 Method

2.2.1 Participants

Across five studies, a total of 361 volunteers were recruited. Four participants were subsequently removed from analysis for not completing the experiment or for reporting having vision that was not normal or corrected-to-normal, leaving a total of 357 participants in the final sample. For a breakdown of demographics, see the method section specific for each study below.

Three of the studies (lines, images, mono-animate) recruited participants over the internet using Amazon Mechanical Turk (i.e., Mturk). These participants were compensated financially for their time. The other two studies (stereo-animate, real objects) recruited participants into the lab. These participants were recruited from the Princeton community and either received course credit or financial compensation for

their time. All participants were briefed about the task before volunteering and were debriefed on the purpose afterwards.

2.2.2 Defining shape continuity

Twenty shape-families of novel objects were created for this project; see **Figure 4**. Each shape-family consists of the same object at different levels of continuity: objects within a shape-family are essentially identical except for the difference in continuity in one area of its shape. Across shape-families, objects varied substantially to best test the robustness and generalizability of continuity in the curvature bias. Because each shape-family tests different pairings of continuity, there were 60 objects tested.

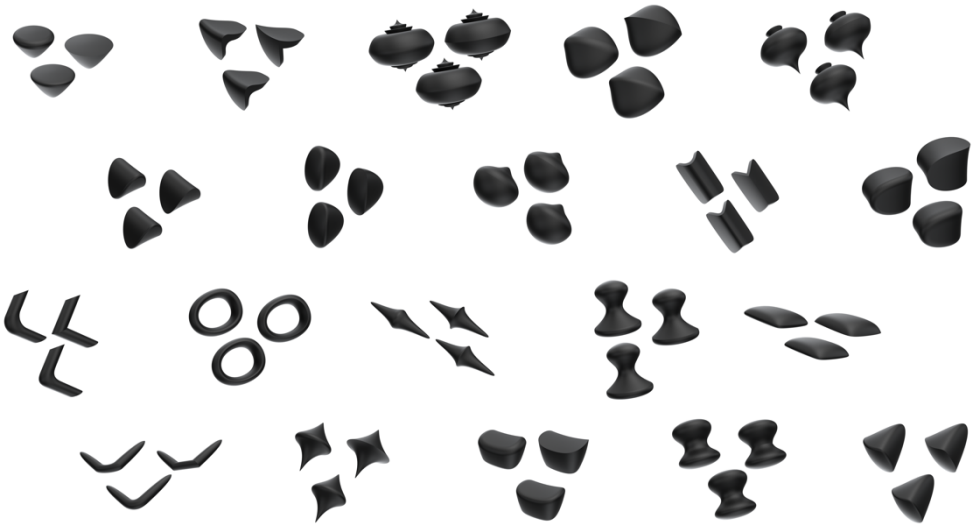


Figure 4. The stimuli consist of 20 shape-families, each containing an example of POS, TAN, and CUR. Here they are presented as images.

We created novel objects, designed to be non-representational to avoid the possible effects of semantic associations on preferences (Vessel & Rubin, 2010). The

objects, in other words, were abstract shapes. Each object represents one connected manifold constructed by joining two to four surfaces along their boundaries. A single transition between two of those surfaces was the primary manipulation examined in this paper. The transition either featured POS, TAN, or CUR. If a shape had more than one transition, the other transitions always featured POS. Specifying TAN and CUR creates a transitional surface by transforming a portion of the two joining surfaces; the width of the transition was held constant across a shape-family.

The objects were created using a 3D modeling software, Autodesk Alias (version 2017). This software is used by industrial designers because it allows for strict control over surface geometry and smoothness (Kahney, 2013, p. 118). Within the software, surfaces were constructed using Non-uniform rational basis splines (NURBS), which are comparable to the aforementioned Bézier curves in the context of continuity. Where multiple surfaces meet, the software allows for manipulation over the type of transition desired. Continuity can therefore be defined by following:

POS: $f\ (\mathrm{x}) = g\ (\mathrm{x})$

TAN: $f^{1}\ (\mathrm{x}) = g^{1}\ (\mathrm{x})$

CUR: $f^{2+}\ (\mathrm{x}) = g^{2+}\ (\mathrm{x})$

The location of a theoretical point (x) on a transition between curves or surfaces is described by two functions, $f(x)$ and $g(x)$; see **Figure 1**. POS occurs when the two surfaces are joined and there is no derivative at the point where they meet. In other words, there is no transitioning surface; there is only an edge. TAN occurs when the two surfaces are joined, and a transitioning surface is created. Points along the transition

represent $f(x)$ and $g(x)$ meeting at the first derivative, but no higher. If $f(x)$ and $g(x)$ meet at the second, or higher derivative, the transition is categorized as CUR.

The presentation of stimuli was comparable across the five studies, but they were not exactly the same. Additional details regarding the stimuli are in the following method sections specific to each study.

2.2.3 Procedure

2.2.3.1 General procedure

Across five different studies, five different modes of visual presentation increasing in "realness" of objects were tested: lines, images, mono-animate, stereo-animate, and real objects. Across studies, objects were presented as dark-gray or black in color, and the background was light-gray or white. The size of the stimuli varied between studies, and there was no control over the size of their presentation in the online studies. For the lab studies, the size the objects was consistent across participants.

In all studies, to test the curvature effects, pairs of objects within a shape-family were presented side-by-side and participants were asked, "Which do you prefer?" The two objects differed only in continuity along one boundary. Participants were instructed that the pairs of objects may appear similar but were indeed different. They were not provided information regarding how they differed. The order of pairings was randomized, as well as the side each object appeared on. There was no time-limit on responses; once participants selected the object they preferred; the next trial was presented. Every participant rated every pair of objects within every shape-family for a total of 60 trials.

Following the main aesthetic preference task, visual discriminability of continuity was tested. Four objects from the same shape-family (POS, TAN, CUR, CUR) were presented and participants were asked: “Which two are the same?” Since the duplicate object was always CUR, the correct answer required selecting the two objects featuring CUR. Performance on this task is referred to as sensitivity and it is a measure of people’s ability to discriminate subtle differences among stimuli. Because there were 20 shape-families, with no repeats, the second part consisted of 20 trials.

2.2.3.2 Stimulus presentation

In each of the five studies, a different mode of presentation was used: lines, images, mono-animate, stereo-animate, and real objects. Line, image, mono-animate were online studies developed using the jsPsych (version 6.0.2) platform (de Leeuw, 2015). To promote more engagement in our online studies, we implemented a five-second delay after a stimulus was presented; participants could provide their response any time after the delay. Stereo-animate and real objects were in-lab studies and no delay was implemented.

2.2.3.3 Study-1: Lines

Object boundaries were presented as simple 2-D lines. To generate these stimuli, the contour that differed in continuity was extracted from each object. To do this, the objects were rotated and then 3D geometry was projected onto a 2D plane using the 3D modeling software. Objects belonging to the same shape-family were rotated exactly the same way. The contours were then cropped to the region where the transition occurs. The

crop transformed the closed curves into open ones. This was necessary to generalize our results to prior literature studying the impressions of line-drawings (e.g., Guthrie & Wiener, 1966; Lundholm, 1921; Poffenberger & Barrows, 1924; Uher, 1991). The stimuli were presented as black lines on a white background, see **Figure 5**.

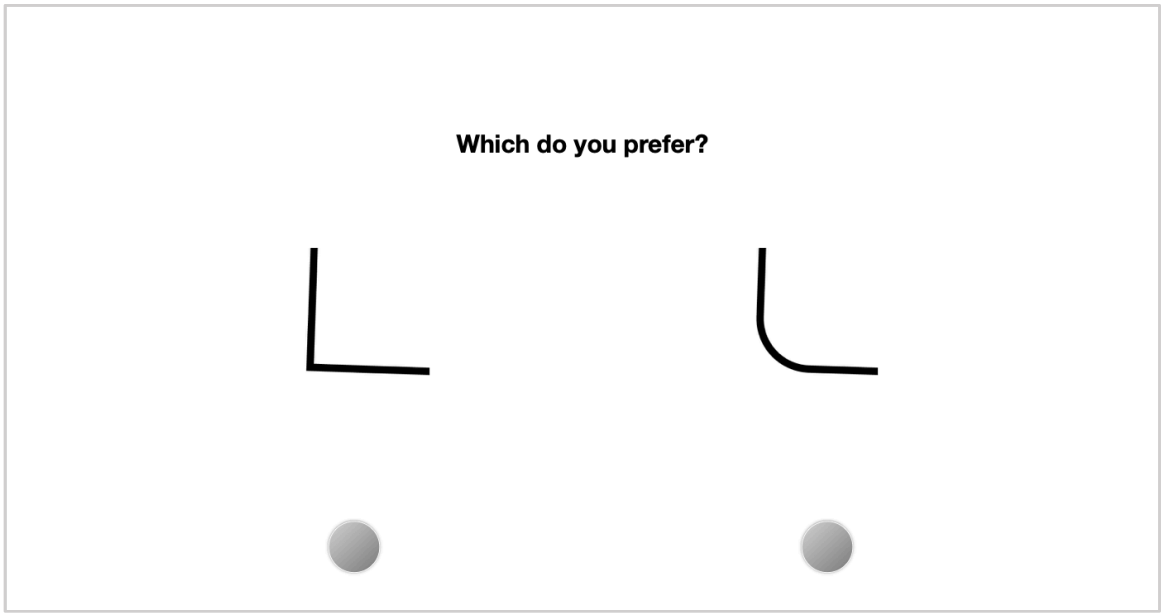


Figure 5. Example of a trial in the study testing continuity presented as lines. The example on the left is POS, right is TAN. The circles below are buttons users clicked on to make their selection.

Data from 99 participants (mean age = 40.55, *SD* = 13.54; 59 female, 50 male) were analyzed from this study. On average, the study was completed in 13.27 minutes (*SD* = 6.56).

2.2.3.4 Study-2: Images

Objects were presented as static images. The objects were first rotated in the 3D modeling software so that the region where the continuity was manipulated was visible to a virtual camera. Objects belonging to the same shape-family were rotated the exact same

way. The images were rendered by a virtual camera using V-Ray (version 3.6) for Rhino (version 6). V-Ray is software that simulates physical light to produce realistic 2D images of 3D objects. The stimuli were presented as black objects on a white background; see **Figure 6**.

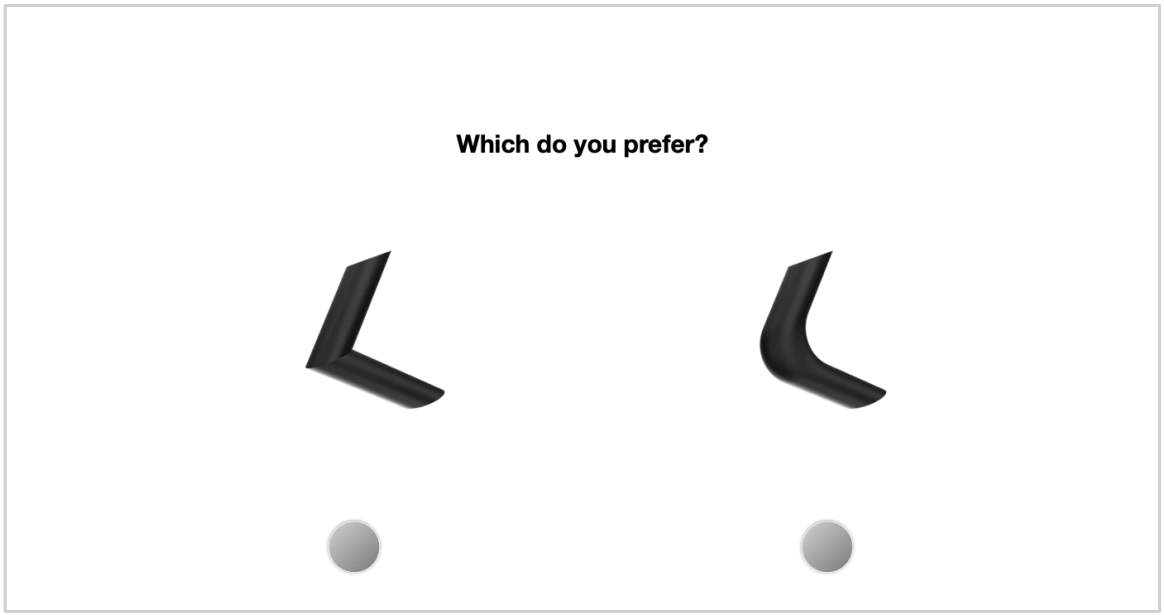


Figure 6. Example of a trial in the study testing continuity presented as images. The example on the left is POS, right is TAN. The circles below are buttons users clicked on to make their selection.

The renderer simulated a plastic material for the objects. Specifically, the diffuse color of the shader was an opaque black (R = 0.03, G = 0.03, B = 0.03); roughness = 0. The reflection color was white (R = 0.94, G = 0.94, B = 0.94); reflection glossiness = 0.65, Fresnel effects enabled, reflection index of refraction = 1.5. The bidirectional reflection distribution function was left using the default settings (type = Blinn, anisotropy = 0, rotation = 0, derivation = local axis Z.). The refraction color was black (R = 0.00, G = 0.00, B = 0.00); refraction glossiness = 1, index of refraction = 1.6. The fog color was white (R = 1.00, G = 1.00, B = 1.00); fog multiplier = 1, fog bias = 0. The

objects featured a subtle texture due to a bump map created from black and white Perlin noise (amplitude = 1; frequency = 9; amplitude ratio = 0.4; frequency ratio = 2; persistence = 1; octaves = 3); bump multiplier = 0.03.

The following are the settings for the camera in V-Ray: F-number = 8, ISO = 625, shutter speed = 100. Primary illumination was calculated using the brute force method, while secondary rays were calculated using light cache (subdivs = 1500, sample size = 0.02, raytrace = 4). Raytracing settings were as follows: noise limit = 0.01, min subdivs = 1, max subdivs = 18, shading rate = 6, bucket size = 32, Lanczos antialiasing filter = 2, adaptive lights = 16, max trace depth = 5, max ray intensity = 20, opacity depth = 50. The objects were rendered in the default Rhino (version 6) spherically mapped environment.

Data from 98 participants (mean age = 36.26, SD = 11.99; 48 female, 48 male, 1 non-binary, 1 undisclosed) were analyzed from this study. On average, the study was completed in 13.76 minutes, *SD* = 4.75.

2.2.3.5 Study-3: Animations (mono-animate)

Objects were presented as monoscopic animations (mono-animate). Images can provide only limited information about an object's shape. Displaying the objects as rotating animations provides an enhanced sense of an object's 3D geometry (Kourtzi & Shiffrar, 1999). To present the stimuli as mono-animate, the 3D objects were converted from NURBs to polygons and exported as OBJ files; tessellation tolerance = 0.01. The 3D object files were then rendered on the participant's screen using WebGL by implementing the p5 (version 0.7.3) JavaScript library. The stimuli were presented as black objects on a light gray background; see **Figure 7**.

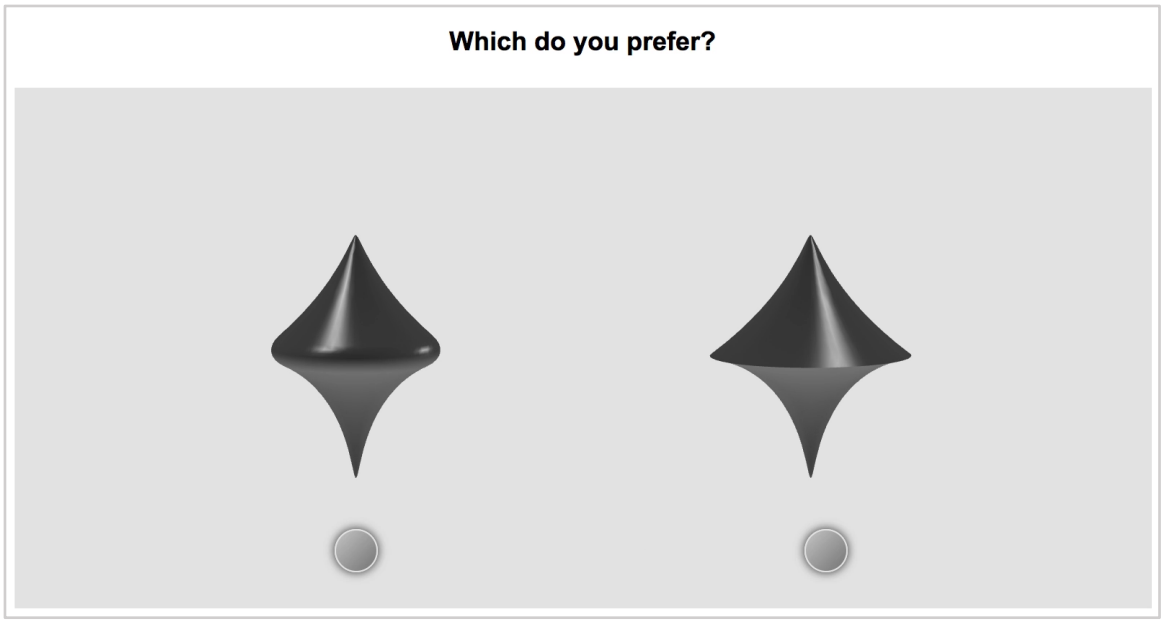


Figure 7. Example of a trial in the study testing monoscopically presented animations. The example on the left is TAN, right is POS. The circles below are buttons users clicked on to make their selection.

The rendering settings of p5, more specifically, were as follows: specularMaterial = 80; normalMaterial = enabled; background = 230. The objects were illuminated using the following settings: ambientLight = 150, pointLight = (250, 250, 250, 1000, 600, 0), pointLight = (250, 250, 250, -1000, 600, 0), directionalLight = (250, 250, 250, 0, 800, 0). These settings illuminated the objects using ambient light, a directional light directly overhead, and two overhead spotlights positioned to the left and right.

The objects rotated around their locally defined origin at approximately 36 degrees per second; the exact number depended on the participant computer's computational power. The object's rotation was randomly moving around its local Z-axis, Y-axis, and X-axis. The animated rotation was intended to provide additional information about the 3D surfaces without necessitating participants to manipulate the objects.

Data from 99 (mean age = 35.45, SD = 11.33; 55 female, 44 male) participants were analyzed from this study. On average, the study was completed in 16.54 minutes, *SD* = 6.11.

2.2.3.6 Study-4: Virtual objects (stereo-animate)

Objects were presented as stereoscopic animations (stereo-animate) using virtual reality technology. This study, relative to mono-animate, provides the additional information that comes with stereoscopic vision. To present the stimuli as stereo-animate, the same OBJ files used in the mono-animate study were presented using virtual reality (VR) hardware. The stimuli were presented as black objects on a light gray background; see **Figure 8**.

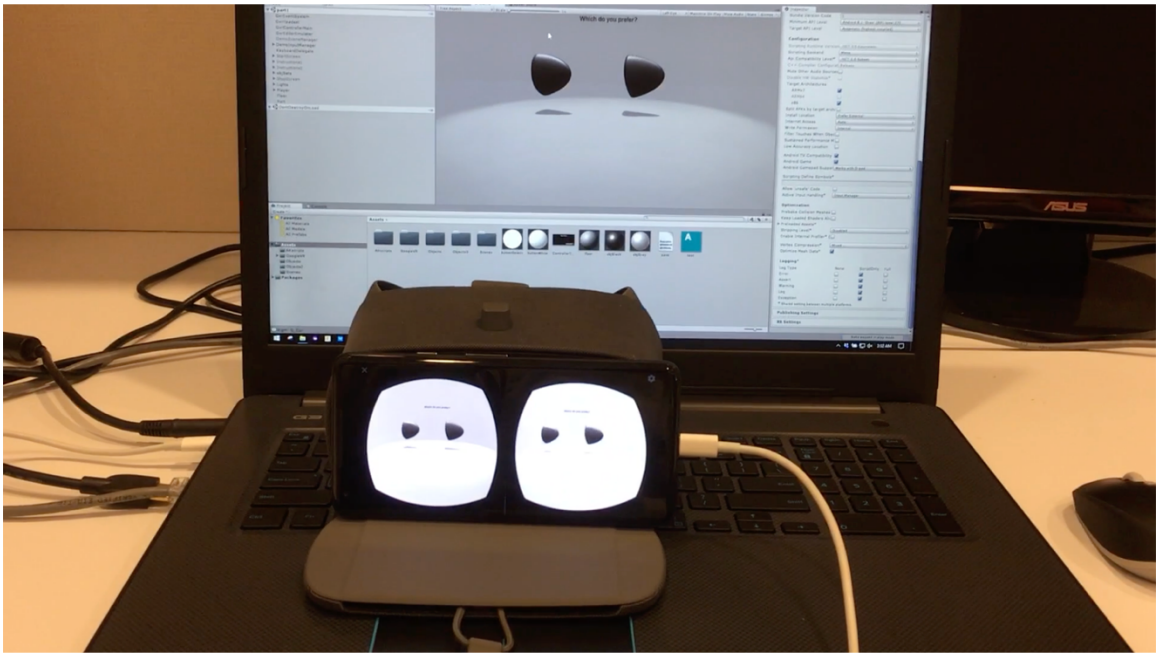


Figure 8. Example of the stereoscopic animations. The computer is displaying the Unity platform and is connected to the VR headset displaying the same animation stereoscopically. The headset is worn over the head so that each eye is presented with only one image.

The 3D models were presented on a Google Pixel 2 smartphone housed within a Google Daydream View (version 2017) VR headset. The VR headset contains two lenses which project slightly different images to each eye; otherwise known as stereoscopy. The different images provide information about a shape from slightly different angles. Viewed in this mode, the dual images create binocular disparity and provide depth information. The VR headset also tracks head movements in virtual space; participants could move their head side to side and see slightly around the objects.

Participants engaged in the first and second task by using a controller to select objects. To make a selection, participants pointed towards an object and then pressed a button on their controller.

This study was programmed using Unity (version 2018.2) with the Google VR SDK (version 1.180). Rendering settings for the VR scene was manipulated so that the objects were perceptually similar to the other studies. The object color appeared dark gray (R = 191, G = 191, B = 191). Other settings are as follows: metallic = 0.5, roughness = 0.5, specular highlights = enabled. The floor of the virtual space was white, metallic = 0, roughness = 0.41, specular highlights enabled; reflections enabled. Beyond the floor, the infinite space of the environment was gray (R = 71, G = 71, B = 71). The objects were illuminated by a directional-light and two spotlights. The directional-light was positioned six units above the objects; white, mode = baked, intensity = 0.4, indirect multiplier = 1.1, shadow type = soft shadows. The spotlights were placed six units above and four units to the left/right of the objects. The spotlights were white; range = 10, spot angle = 80, mode = baked, intensity = 1.1, indirect multiplier = 1.1, shadows = disabled.

The objects randomly rotated around their local origin at a rate of 36 degrees per second; fps = 60. The scene and stimuli appeared similar to the other studies.

Data from 32 participants (mean age = 19.77[2], *SD* = 2.89; 20 female, 12 male) were analyzed from this study. On average, the study was completed in 15.35 minutes, *SD* = 5.63.

2.2.3.7 Study-5: Real objects

Objects were real, 3D-printed objects. The original NURBS models were converted to STL files and then 3D printed; tolerance = 0.01, max edge length = 0.04. The objects were printed using the HP Multi Jet Fusion in Nylon Plastic PA12; print layers = 0.08 mm, tolerance accuracy = 0.25 mm. Each object was sized to fit within a bounding box of 2 x 2 x 1.5 inches. The stimuli were presented as dark-gray objects on a light-gray background, see **Figure 9**.

[2] Because the participants were recruited from the local community, people over the age of 35 did not have to disclose their age to protect their privacy. This statistic omits the age of two participants for this reason.



Figure 9. Example from the participant's perspective in the study testing real objects. A researcher would sit across from them and place the objects during each trial. Participants could touch and manipulate the objects in this study.

The objects had a subtle texture that refracted light. They were presented on a light-gray polyester cloth that covered an elevated round platform (diameter = 12 inches). The platform was centered on a round table covered in a black cotton cloth. The participant and researcher sat on either side of the table. Illumination was provided by a diffused light overhead and two spotlights. The diffused light was an A19 LED; positioned 32 inches above the objects and diffused using a translucent white photography umbrella; lumens = 800, color = 5000K. The two spotlights were BR30 LEDs; they were positioned 15 inches above the objects and 14 inches to the left/right; lumens = 650, color = 5000K. The setup was visually similar to the other studies.

To run this experiment, a researcher physically presented the objects during each trial. Computerized instructions facilitated randomization. The objects were stored in

coded drawers and the instructions told the researcher which drawers to access and where to place the objects. The setup was designed to make it difficult for a researcher to influence the outcome of a trial because the conditions were always obfuscated.

Both the participant and the researcher wore black nitrile gloves to keep the objects clean, numb mechanoreceptor sensitivity to texture, and control the amount of light reflected by hands. After a set of objects was laid down in front of a participant, they could pick them up and viewed them from all sides. Whenever the participants were ready to provide a response, they pointed and/or vocalized their selection.

Data from 30 participants (mean age = 22.27, *SD* = 6.92; 21 female, 9 male) were analyzed from this study. On average, the study was completed in 44.70 minutes, *SD* = 13.52. This study took longer to complete because the researcher had to physically place each object.

2.2.4 Analyses

Data were analyzed using R (version 3.6.0). We first describe the proprotions preferring the curvier and smoother continuity while also reporting the effect size (Cohen's *d*) when compared to random chance (*mu* = .50). Generalized models were analyzed using the 'lme4' package (version 1.1-21) and are specified in this chapter using that package's syntax.

The primary measure was the likelihood that one object was picked over the others. To test if there was a significant bias present, we analyze the base log odds (β_0) in a mixed effect model. This analysis provides a method for isolating variance attributable to aesthetics while separating it from the random effects each participant and study brings

into a sample (Corradi et al., 2019). Specifically, we analyzed preferences between pairs by running a generalized linear mixed model with a binomial logit link function; $C \sim (1 \mid participant) + (1 \mid study)$; C is a binomial value representing whether or not the object with higher continuity was picked in a trial (POS < TAN, POS < CUR, TAN < CUR); variance attributable to each *participant* (u_j) and *study* (u_k) are accounted for as random effects.

As a secondary analysis, we also report planned comparisons by reporting the results of separate models for each pairing of different types of continuity. The results of the planned comparisons describe the likelihood TAN is preferred over POS, CUR over POS, and CUR over TAN. The reason we report this as a secondary analysis is because of our pre-registration; prior research on the curvature effect could only inform a hypothesis that conflates TAN and CUR; therefore, these planned comparisons were not our primary analysis. Note that later in the results section, we also report the participant related variance attributable to the different pairings to describe potential patterns by which preferences deviate from the curvature effect. Among industrial designers[3], there is a debate contending that preferences between POS v. TAN, or POS v. CUR, are more idiosyncratic than between TAN v. CUR.

The second part of each study examined participants' sensitivity to the differences between continuity; specifically, TAN v. CUR since their difference is subtle. We hypothesized that people's performance would significantly affect the likelihood that they

[3] Based on one of the author's, A. Kurosu, personal experience and conversations with other industrial designers. The notion of this debate was pre-registered.

preferred higher continuity (TAN < CUR). We ran a modified analysis; adding a measure of *sensitivity* (β_1) to our model, $C \sim$ sensitivity + (1 | participant) + (1 | study); in this analysis, *C* specifically refers to the aesthetic of CUR over TAN. We predicted that *sensitivity* would be a significant predictor of greater preference for CUR.

2.2.5 Pre-registration

Pre-registration of our sample sizes, hypotheses and any deviations[4] can be found online, https://osf.io/z9xf6. Note that in the registration '2D-1' is the study with lines, '2D-2' is the study with images, '3D' is the study with mono-animations, 'VR' is the study with stereo-animations, and 'RL' is the study with real objects. The change in labeling was made to improve the clarity of the chapter.

[4] deviations will be published after the manuscript is submitted for peer-review.

2.3 Results

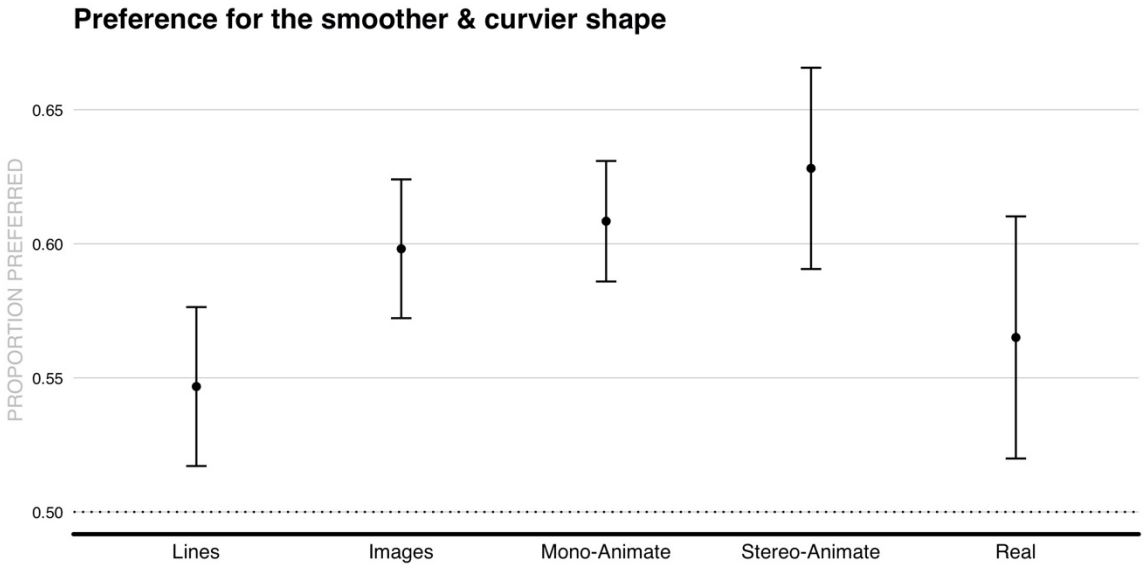


Figure 10. Proportion TAN or CUR (smoother and curvier shape) preferred over POS. The dotted horizontal line at .50 represents random chance. Higher numbers represent stronger evidence for the curvature effect. Error bars represent the 95% confidence intervals based on standard error.

As predicted, the results demonstrate an aesthetic bias for smoother and curvier shapes. People are significantly more likely to prefer an object featuring TAN or CUR over POS; see **Figure 10.** This occurs regardless of other undefined differences between shape-families. It also occurs regardless of differences between modes of visual presentation. The effect size and proportion preferring TAN or CUR over POS for each study are as follows: lines, $M = .55$, $d = .180$; images, $M = .60$, $d = .433$; mono-animate study, $M = .61$, $d = .549$; stereo-animate study, $M = .63$, $d = .682$; real objects, $M = .57$, $d = .298$. Although the strength of the effect varies across visual modes, the results of each study demonstrate the curvature effect.

Compiling the data across studies, we examined preferences between different pairs of continuity types. Comparing POS v. TAN, preferences for objects featuring TAN

were significantly more likely, $\beta_0 = .573$, $p < .001$. Comparing POS v. CUR, preferences for CUR were significantly more likely, $\beta_0 = .669$, $p < .001$. Comparing TAN v. CUR, preferences for CUR were still significantly more likely, $\beta_0 = .151$, $p = .008$. Taken together, the results of these analyses support our hypothesis that greater continuity is more likely to be preferred.

To describe consensus among preferences, we analyzed the participant variance. Participant variance represents idiosyncratic deviations from the curvature effect. The results show that preferences for shape continuity are not necessarily indicative of every trial. It appears to be strongest in trials comparing TAN v. CUR, then between POS v. TAN, or POS v. CUR. Between TAN v. CUR, participant variance was very low, $u_j = .117$. Comparing POS v. TAN and POS v. CUR, participant variance was higher; $u_j = 1.748$, and $u_j = 1.495$ respectively. Thus, our results provide supporting evidence of the curvature effect, however an analysis of participant variance shows idiosyncratic deviance is higher when POS is considered against TAN or CUR.

The second part of each study tested people's sensitivity to different types of continuity. We hypothesized that greater sensitivity predicts a greater likelihood for preferring CUR over TAN. As hypothesized, the data show that sensitivity bears a significant effect on preferences, $\beta_1 = .142$, $p = .006$. People who were more sensitive to differences in continuity were more likely to prefer stimuli featuring greater continuity.

2.4 Discussion

Our five studies suggest that people prefer smoother and curvier shapes; i.e., the curvature effect. While the current literature already provides robust evidence for the curvature effect, what remained unknown was how to objectively define this effect and whether the effect generalizes across different modes of presentation. Given the different stimuli and contexts that may be related (e.g., Gómez-Puerto et al., 2016), an appropriate definition needed to be generalizable to shapes presented in a variety of ways. We contend that 'continuity' is the visual information that stitches prior findings together. Our results show the curvature effect using an objective definition generalizing to stimuli ranging from lines to real objects.

Based on our results, we contend that the curvature effect is an aesthetic bias for greater shape continuity. This objective definition brings a framework for generating new questions. For example, this study tested three categories of continuity, but not their mixtures. Previous research has considered mixtures of what was thought to be only two categories (e.g., Bar & Neta, 2006; Lundholm, 1921). Framing the curvature effect in terms of three categories of continuity therefore highlights a need to revisit prior findings. For example, how do different mixtures of POS, TAN, and CUR affect aesthetics? Coincidently, this question has already been discussed by industrial designers, yet it remains scientifically untested (e.g., Holland, 2009).

It is important to understand that our data do not suggest that people *always* prefer greater continuity. Our analysis of participant variance shows that deviance from the curvature effect is greatest among trials comparing POS v. TAN, and POS v. CUR,

relative to TAN v. CUR. One explanation is that jump from POS to TAN represents a larger qualitative difference than going from TAN to CUR. Indeed, this finding is consistent with previous research showing that the curvature effect does not predict preferences in every situation (Corradi et al., 2019). What may explain idiosyncratic preferences could be factors such as emotional processing (Blazhenkova & Kumar, 2018), expertise (Corradi et al., 2019; Cotter et al., 2017; Ruta et al., 2019; Silvia & Barona, 2009; Vartanian et al., 2019), and personality (Banaei et al., 2019; Cotter et al., 2017). Future research revisiting these measures of individual differences in light of this new definition of the curvature effect could bring greater specificity to this literature.

The value of testing the aesthetics of continuity is not only theoretical, but also has applied value. Industrial designers have long obsessed over the curvature of products (e.g., Mickle, 2019). Surfaces with greater continuity are considered Class-A[5] geometry and are aesthetically valued (G. Holland, 2009). By using the nomenclature of continuity to objectively define the curvature effect, this study directly assesses an old industrial design heuristic for the first time. Furthermore, this research has the potential to impact the design of everyday life by moving design heuristics into a design science (e.g., Kurosu, 2019).

---

[5] The origin of the terminology is unclear. One possibility is Doug Satzger, a designer at Apple, who referred to the highest quality surfaces of a product being the front. B surfaces are the sides, C surfaces are the back, and D surfaces are the insides (Kahney, 2013, p. 208). Another possibility is attributed to Mercedes-Benz, where 'A' stood for the German word "Aussenhaut" which translates to 'outside surface,' and it may tied to the perception of monotone curvature and torsion (G. Farin, 2006).

This is the first scientific study supporting the notion of preference for Class-A surfaces. Considering the wealth traded between countries in the form of designer goods[6], there is an enormous economic value in understanding what makes products enticing. Take for example the 2012 landmark court decision to award Apple $1B in part because their design patent on a rounded rectangle was infringed upon (Wingfield, 2012). While some mocked the absurdity of patenting such a simple shape (Bonnington, 2012), designers felt differently (Ryan, 2016). What therefore may be missing in the public discourse is the science on shape continuity's aesthetic.

[6] The G20 exports $200 Bilion annually. This number is based on the number of "Design" creative goods exported by the G20 over the last five recorded dates (2011, 2012, 2013, 2014, 2015) from the united nations conference on trade and development (UNCTAD); unctadstat.unctad.org. Design includes the following product categories: architecture, fashion, glassware, interior, jewelry, toys. Data about the G20 includes the following countries: Argentina, Australia, Brazil, Canada, China, France, Germany, India, Indonesia, Italy, Japan, Korea, Mexico, Russia, Saudi Arabia, South Africa, Turkey, United Kingdom, United States.

3.

Reliability of object impressions.[7]

3.1 Introduction

People constantly encounter objects they haven't seen before. Despite the frequency of these events, it remains unclear how people form impressions of novel objects. Whether it be window shopping or gallivanting through a museum, without much thought, people can easily dismiss most objects in view and gravitate to just a few. One possibility is that the shape of these objects affects their nascent impression. From prior research on perceptual categorization, it is known that people are meaningfully sensitive to nuanced shape-differences (e.g., Folstein et al., 2013; Freedman et al., 2003; Goldstone & Steyvers, 2001). Yet few publications have directly examined the effects of shapes on impressions, and even fewer using novel and abstract objects; some examples being: rectangular width-length ratio (e.g., McManus, 1980), radial frequency (Y.-C. Chen et al.,

[7] Work based on a published manuscript by Kurosu, A. & Todorov, A., *Journal of Vision,* 2017.

2016), and curviness (Amir et al., 2011; Bar & Neta, 2006; Marco Bertamini et al., 2016; Gómez-Puerto et al., 2016). The limited differences in which shapes have been explored in prior research could be due to two reasons: 1) a lack of algorithms for systematically manipulating the shape of novel objects, 2) difficulty defining meaningful metrics by which to manipulate shapes (Phillips et al., 2010; Wilkinson et al., 1998). For example, "curviness" is sometimes operationalized as the opposite of "straight", while other times as "sharp" or "angular" (Gómez-Puerto et al., 2016). Cognizant of these issues, we introduce a new technique for generating stimuli and revisit the general question of whether shapes influence an object's impression without preconceptions of what constitutes meaningful shape-differences.

Studying the effects of an object's shape on impressions is difficult, because the moment a person knows that an object is present, semantic information is already available (Grill-Spector & Kanwisher, 2005). To focus on the possible effects of shapes themselves, we test abstract objects because any associations to them are presumably random across the population. Crucially, these objects are unique in that they also maintain parametric relationships across their differing shapes. Thus, if shapes have a systematic and knowable role in the formation of a novel object's impression, we should be able to observe shape-differences being isomorphic with impression-differences.

We test the reliability of shape's effect on impressions by analyzing consistency and consensus. Consistency (i.e., intra-rater agreement, test-retest reliability, personal taste) signifies the extent to which individuals form the same impression upon re-evaluation. Consensus (i.e., cross-observer agreement, inter-rater reliability, shared taste, intersubjectivity) signifies the extent to which impressions are similar across people.

Statistically, consistency is a pre-condition for consensus. Conceptually, consistency captures the reliability of the subjective experience of stimuli and consensus captures the reliability of the experience being shared across the population. Consistency and consensus can be used to model the contributions of subjective (i.e., private taste, individual preference) and objective (i.e., shared taste, common preference) evaluations to impressions (Germine et al., 2015; Hönekopp, 2006; Martinez et al., 2020).

To produce our stimuli, we manipulated 3D shapes through parametric algorithms using a method called 'procedural generation' (for an overview: Krispel et al., 2015). A single parametric algorithm generates a 'shape-family' of objects per inputted values. The resulting objects share relatable shapes. Just as how genetic siblings appear relatively similar and yet different, objects belonging to the same shape-family appear similar and yet different. While there have been previously published novel object families such as "greebles" (Gauthier et al., 1998) and "fribbles" (Williams & Simons, 2000), the shape-families we introduce here are different in that they can be defined within a continuous parametric space. These objects therefore afford a unique opportunity to test if the impressions of complex novel objects are predictable given the parametric values defining their shape.

We also test impressions of faces to situate the reliability of object-impressions. People form consistent impressions of faces, and different people largely agree (i.e., consensus) on these impressions (for a review: Todorov et al., 2015). First impressions of faces can be codified in terms of how trustworthy (i.e., valence of evaluation) and dominant they seem (Oosterhof & Todorov, 2008; Todorov et al., 2008), as well as how attractive they are (Sutherland et al., 2013). Evaluations of trustworthiness and

dominance are related to both judgments of approachability and danger (Todorov et al., 2008; Todorov & Duchaine, 2008).

In contrast to the literature on impressions of faces, the literature on impressions of novel objects is fragmented and unclear. For example, Duckworth and colleagues (2002) observed that people consistently reacted to approach or avoid novel stimuli. However, recent studies have not observed strong consensus amongst preferences for novel and abstract images (Rodway et al., 2016; Vessel & Rubin, 2010). Taken together, prior literature makes it uncertain whether shapes reliably affect impressions of novel objects. We build upon prior research by measuring four impressions that are applicable to both faces and novel objects: approachable, beautiful, dangerous, and likable. Approachable and dangerous relate to approach and avoidance behavior. Beautiful and likable relate to aesthetics and preferences.

In study-1, we first verify the possibility that different people can independently come to share the same impressions of novel objects. We do so by probing the reliability of impressions in terms of consistency and consensus. Conceptually, this phenomenon would be similar to when random people share impressions of strangers. Therefore, pictures of faces were included into the design of study-1 to compare the reliability of impressions from novel objects versus people. After verifying that people can share impressions of novel objects, it was then possible to test if the objects' shapes are a contributing factor. To the extent that consensus forms, we expected the underlying shape-parameters to be predictive of impressions. The results of study-1 confirmed that shapes affect impressions. However, a greater variety of shapes should have been tested. Based on study-1 alone, it could not be ascertained whether consensus emerges only for

some evaluations like dangerousness or if it emerges when specific shapes maps to specific evaluations. In study-2, we therefore included two additional shape-families and focused on impressions of beauty and dangerousness because they yielded the lowest and highest levels of consensus in study-1. We observed that consensus varied as a function of both, the type of evaluation and the type of shape-family. Together, these studies demonstrate that people can share impressions of objects they haven't seen before when the shape of the object maps to a specific evaluation.

3.2 Study-1

3.2.1 Introduction

The purpose of this first set of studies is to test the reliability of people's impressions of objects. Because the only differences between objects in our studies are their shape, if impressions prove to be reliable, the primary hypothesis is that changes in shape covary with changes in impression. To test this, we use a method for generating stimuli called procedural generation (Krispel et al., 2015). In this method, the shape of objects is defined algorithmically and differences between objects are dictated by parametric variables set within the script. All objects generated from the same algorithm are called shape-families. Objects which are generated from the same shape-family, like siblings, share more commonalities than objects generated from different shape-families.

We started with a single family of novel-objects, referred to as *family-x*. First, we tested whether participants would evaluate the novel stimuli consistently across time. To assess this, we asked participants to report their impression of each stimulus 10 times (the

stimuli were presented in 10 different blocks). If participants are consistent, their evaluations will be positively correlated. Moreover, these correlations should be detectable in the first two blocks of evaluation. We also expected that the first evaluations would predict the last evaluations. To the extent that these hypotheses are confirmed, the consistency in evaluation could be attributed to the effect of shapes on first impressions rather than to an effect of learning the shapes over the course of the experiment. Second, we tested whether consensus would emerge across the four types of evaluation: approachability, beauty, dangerousness, and likability. We compared impressions of novel objects to impressions of faces. Finally, we expected that the consensus amongst impressions of novel-objects would be isomorphic with the underlying shape-parameters, which relate each object from *family-x*. In other words, we sought to demonstrate that different people can come to share the same impression of a novel abstract object because of its shape.

3.2.2 Method

3.2.2.1 Participants

188 volunteers from Princeton University and the neighboring community participated in this study. Volunteers from the community were monetarily compensated, while students were compensated with course credit. Amongst the objects, 25 participants were assigned to each of four evaluations. Amongst the faces, the participants were assigned to different evaluations as follows: approachable = 27, beautiful = 19, dangerous = 21, likable = 21; the uneven assignment was due to a notation error in the random

assignment process. All participants had normal or corrected-to-normal vision. None of the participants had seen the stimuli prior to their participation.

3.2.2.2 Object Stimuli

The set of novel objects used in this study, *family-x,* were procedurally generated using an algorithm housing two parameters we refer to as *points* and *flux*. These particular shape-parameters were selected because the resulting objects did not appear like anything in particular (we later verified the assumption of abstractness and novelty; see the method section of study-2). Within this algorithm, points dictate the number of distal projections (3 - 100) and flux dictates the surface acceleration around these projections (ranging from quick, $1/x^{1.00}$, to slow, $1/x^{4.00}$). 66 unique objects were sampled from this family by assigning random values to its parameters; **Figure 11** plots the objects along its parametric values. For further details, refer to the algorithm provided on our lab's website (www.tlab.princeton.edu/databases) where 3D object files and 2D renderings are also made available. *Family-x's* algorithm was written based on a code called "Spatial Deform" by Michael Pryor (http://www.3d-dreaming.com). These objects were generated using Grasshopper (version .9; http://www.grasshopper3d.com) and then rendered as 2D images (800 x 800 pixels) using Rhino (version 5; http://www.rhino3d.com).

All rendering qualities were set constant across all images using the default rendering engine in Rhino. The objects appeared neutral gray on a black background. This appearance was achieved using the default settings of the program. The default shader (i.e., the specifications for simulating a material visually) was white in color with

0% gloss, 0% reflectivity, 0% transparency, and 1.0 index of refraction (IOR). The shader specified no textures. The emission and ambient color were black. Diffuse lighting was enabled; a directional light was positioned above and slightly to the left of the object emitting parallel rays towards the object.

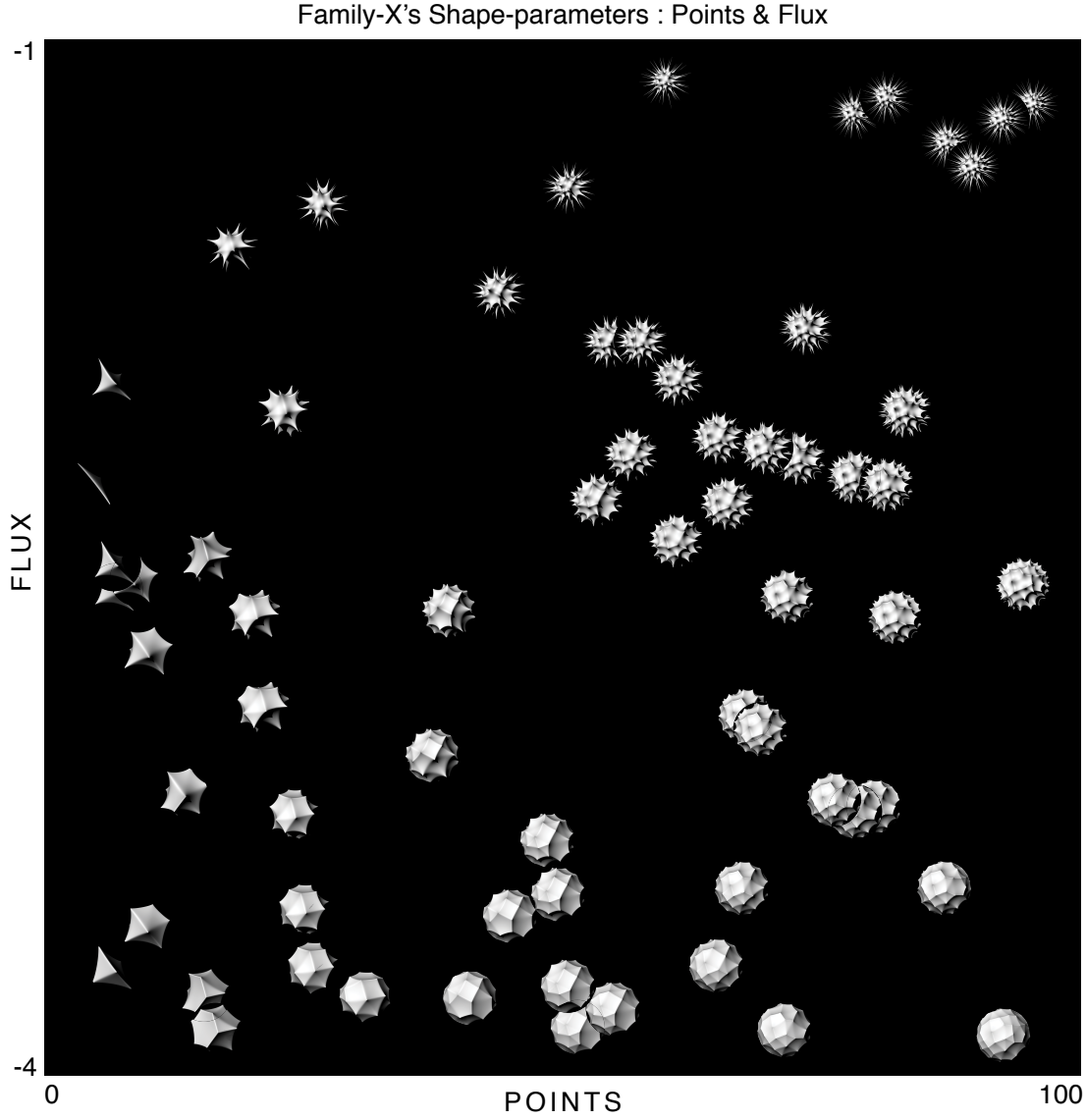


Figure 11. Objects sampled from *family-x*. The stimuli consisted of 66 objects sampled from *family-x*. The shape of these objects is defined by the number of distal points (*points*) and the surface acceleration around these points (*flux*).

3.2.2.3 Face Stimuli

The face stimuli consisted of 66 images from the Karolinska Directed Emotional Faces database (Lundqvist et al., 1998). These images are cropped portraits (562 x 762 pixels) of non-famous Caucasian individuals photographed in front of a gray background with a neutral facial expression.

3.2.2.4 Procedure

Participants sat in a private sound isolated cubicle with a computer. In a 2 x 4 between-subjects design, participants either evaluated objects or faces with one of four types of impressions (approachable, beautiful, dangerous, or likable). For example, one participant would report 'how approachable is this object?' throughout their session.

In each trial, a fixation-point appeared for 500ms in the center of the screen. This fixation-point was replaced by a randomly selected object or face, the evaluation, and a scale ranging from 1 (not at all) through 9 (extremely). Participants were encouraged to go with their 'gut' feeling when keying in their impression. There were no time constraints; objects remained visible until an evaluation was made. Afterwards, a black screen appeared for 500ms. After each image was presented once, constituting a block, participants had an opportunity to take a break. There were ten blocks in this study and it took approximately 30 minutes to complete; see **Figure 12** for a visualization of this procedure. All stimuli were centered upon a black background and presented at a 20° visual angle using CRT monitors. This procedure was coded using PsychoPy version 1.82 (Peirce, 2007).

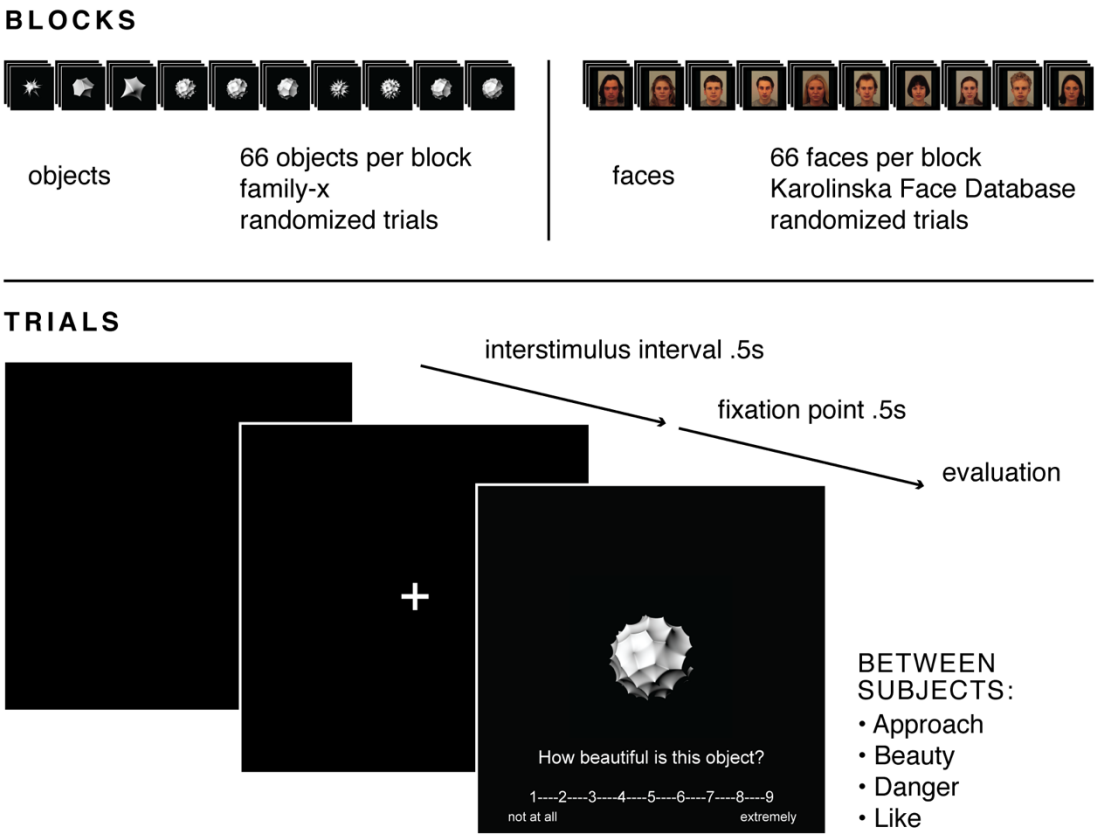


Figure 12. Study-1 experimental design. Participants evaluated 66 stimuli (objects or faces) 10 times using a scale from 1 through 9. Participants either made evaluations of how approachable, beautiful, dangerous, or likable the stimuli seem.

3.2.2.5 Main Analyses

3.2.2.5.1 Consistency

Each item was presented ten times to test how reliable impressions are across time. The 'consistency' was determined by calculating an average Pearson's *r* correlation coefficient per participant. For example, participant-1's impressions from the first-block are correlated with the second-block, the second-block with the third-block, etc.; all possible pairings being correlated together. The resulting correlation coefficients were averaged together to generate a single value representing that individual's consistency.

We then calculated *t*-tests against zero to assess if the level of consistency is significantly greater than zero. Significance testing was calculated using Fisher's *r* to *z* transformed scores.

Because participants rated the objects multiple times, it is possible that our measure of consistency is driven by the effect of learning the stimuli and recalling a former impression. To assess if first impressions are truly meaningful and that consistency is not merely an artifact of learning, we report the correlation between first impressions (in the very first block of stimulus presentation) and every subsequent impression of the same item. Specifically, impressions among the different blocks should not only be highly correlated, but these correlations should be detectable between the very first two blocks, as well as between the first and the last block.

3.2.2.5.2 Consensus

Measures of consensus were calculated using a leave-one-out correlation with the mean statistic (e.g., Engell et al., 2007; Germine et al., 2015). First, the averaged impression, reflective of one participant, is correlated with the average impressions' representative of all other participants. Repeating this process for each participant generates correlation coefficients representing the similarity of one participant's impressions with that of the entire sample. The greater the correlation, the greater the level of consensus. We also calculated consensus using another method where we computed pairwise correlations amongst impressions formed by each participant. Since this second method did not allow for certain tests of significance, we report the results of this alternative method and the results of various significant tests afforded by the

different methods in the appendix. Fisher's *r* to *z* transformed scores were used in *t*-tests against zero to determine if the consensus values observed are significant. We also calculated measures of consensus for the first block only and used the same tests, as outlined above, for significance.

3.2.2.5.3 Shapes

To the extent that there is consensus, we hypothesized that it would occur in a systematic fashion attributable to the constituent shape-parameters. We therefore calculated linear mixed effects regressions to test the extent to which shape-parameters were significantly predictive of impressions. The shape-parameters represent the values used to generate an object. For example, 40 points in *family-x* represents 40 distal extensions of mass in 3D. Not all 40 points would be visible in a 2D rendering, but since the renderer's settings is standardized across objects, a proportionate number of points are represented across renderings. These parametric values were then centralized and standardized per family before being submitted to the regression models. The shape-parameters were tested as fixed effects and the participant-differences were accounted for as a general random effect upon the model's main intercept. In each model, we focused on the proportion of variance, R^2, explained by the fixed factors (Nakagawa & Schielzeth, 2013). A higher proportion of variance attributable to the fixed effects strengthens our evidence that shapes contribute to nascent impressions of objects.

3.2.2.6 Note Regarding Plots

To plot means and standard errors, the statistics were first calculated using the *z* scores and then transformed back into *r* coefficients. The purpose for doing this was to

illustrate the contrasts between different variables while conveying the relative strength of each correlation. However, outside of these plots, untransformed *r* coefficients are reported because they are more conservative. Original and transformed values of consistency and consensus are provided in the appendix.

3.2.3 Results

3.2.3.1 Consistency

Individuals experienced a similar impression every time they were re-exposed to the same item. We first collapsed the level of consistency across the different types of evaluations and calculated separate *t*-tests against zero for objects and faces. The average consistency across ten blocked exposures were significantly correlated for both objects, $t(99) = 24.11$, $p < .001$, $d = 2.41$, $r_M = .66$; and faces, $t(87) = 20.13$, $p < .001$, $d = 2.15$, $r_M = .56$. There were significant differences, however, in the level of consistency across conditions; see **Figure 13.** A between-subjects ANOVA revealed a significant interaction between the type of evaluation and stimuli upon the level of consistency, $F(3, 180) = 7.06$, $p < .001$, $\eta_p^2 = .11$, $\eta^2 = .09$. Consistency significantly differed by evaluation-type for novel-objects, $F(3, 96) = 8.18$, $p < .001$, $\eta^2 = .20$, but not for faces, $F(3, 84) = 1.61$, $p = .19$, $\eta^2 = .05$. Regardless of the differences, significant levels of consistency were found in every condition.

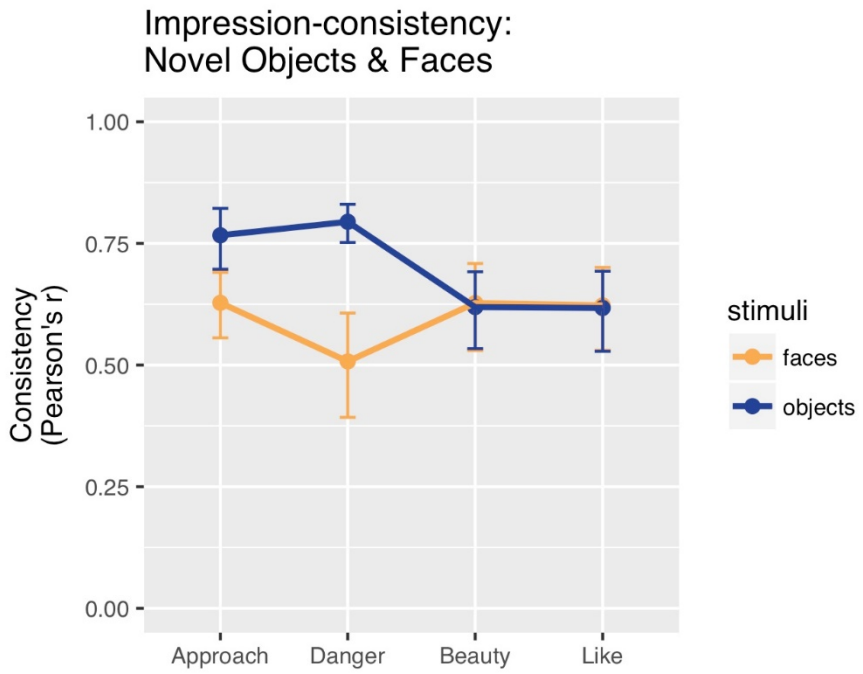


Figure 13. The level of consistency amongst object-impressions was equal-to or greater-than face-impressions. The means are plotted with standard error bars representing the .95 confidence interval.

Were the first impressions meaningful? As shown in Figure 4, the consistency of later impressions with the first impressions remained stable, although in some cases, it did decrease. Examining the correlation between first and second impressions, consistency was significant for faces, $t(87) = 20.70$, $p < .001$, $d = 2.21$, $r_M = .51$; and objects, $t(99) = 25.19$, $p < .001$, $d = 2.52$, $r_M = .66$. Moreover, the correlation between the first and tenth impressions was also significant for faces, $t(99) = 17.62$, $p < .001$, $d = 1.88$, $r_M = .49$; and objects, $t(99) = 18.25$, $p < .001$, $d = 1.83$, $r_M = .63$. These results support our hypothesis that engaging in the evaluation of novel objects is non-arbitrary for an individual insofar as it elicits consistent impressions. Hence, to estimate the reliability of an individual's impressions, it appears that including just one subsequent repetition is sufficient.

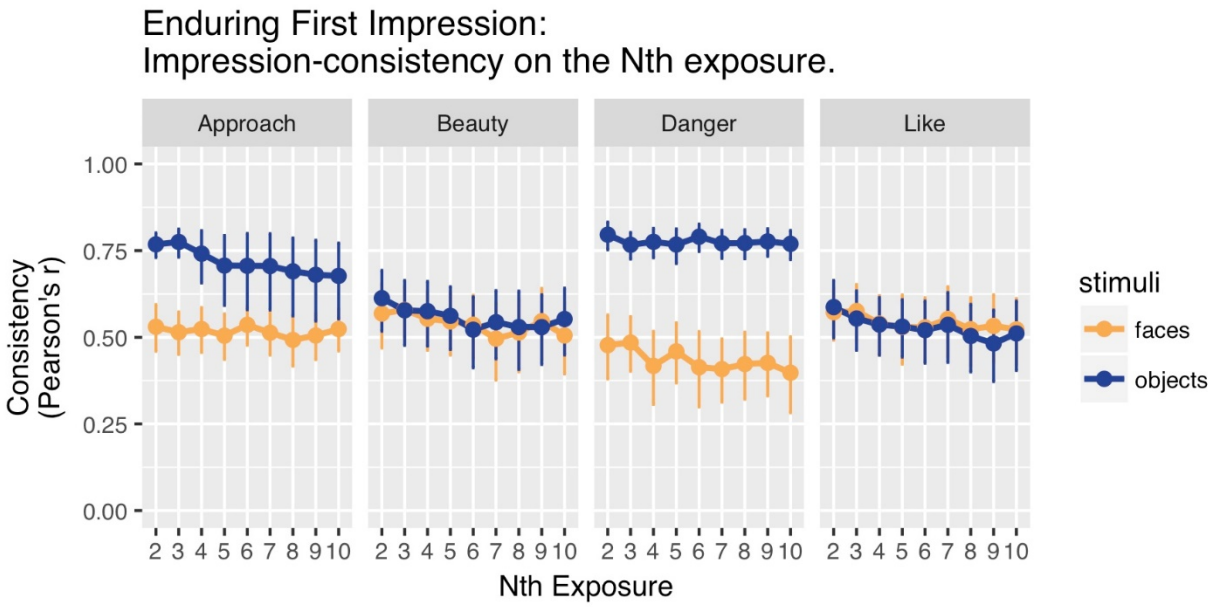


Figure 14. First impressions correlated with the impression on the Nth exposure. The first impressions of faces and novel objects were correlated with their impressions from each subsequent re-evaluation; e.g., first impressions were consistent with the 10th impressions.

3.2.3.2 Consensus

Overall, people appeared to share similar impressions. Using *t*-tests against zero, significant consensus was present for both objects, $t(99) = 10.48$, $p < .001$, $d = 1.05$, $r_M = .55$; and faces, $t(87) = 27.46$, $p < .001$, $d = 2.93$, $r_M = .63$. However, closer inspection revealed large variation in the level of consensus; see **Figure 14**. Consensus for novel

objects was only found for two types of evaluations: approachable and dangerous. This was not the case for faces; consensus was found for all four types of evaluations.

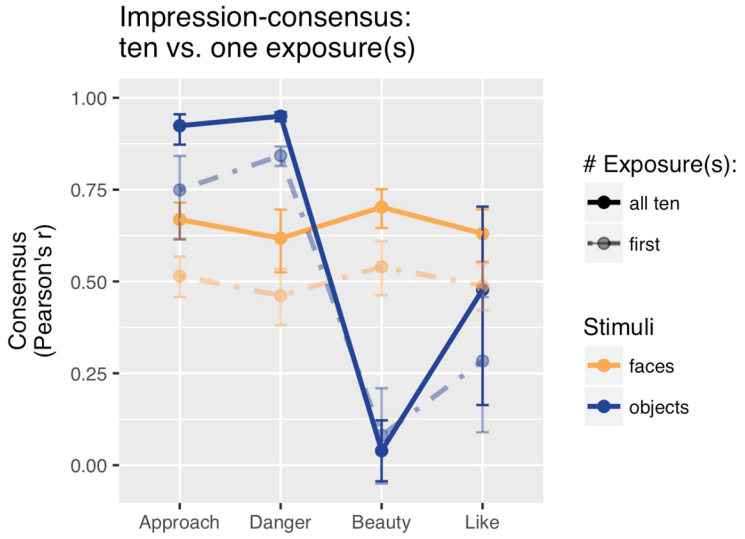


Figure 15. The extent to which impressions are shared varies depending on the type of evaluation and stimuli. The mean consensus values are plotted along with standard error bars representing the .95 confidence limits. Averaging impressions collected across ten exposures (solid lines) results in a higher level of consensus than from first impressions alone (dashed lines), yet the pattern of differences between conditions is unaffected.

Analyzing the different types of stimuli and evaluations together with ANOVA revealed a significant main effect of the type of stimuli, $F(1, 180) = 12.80$, $p < .001$, $\eta_p^2 = .07$, $\eta^2 = .03$; the type of evaluation, $F(3, 180) = 40.42$, $p < .001$, $\eta_p^2 = .41$, $\eta^2 = .29$; and the interaction between different evaluation-types and stimuli, $F(3, 180) = 40.69$, $p < .001$, $\eta_p^2 = .40$, $\eta^2 = .28$. While faces elicited a reliably significant level of consensus across the different evaluation-types, novel-objects elicited highly variable levels of consensus.

The highest level of consensus occurred amongst object impressions of danger, $r_M = .94$, and approach, $r_M = .85$. The difference between these two types of evaluations was not significant, $p = .66$. Following suite was the magnitude of consensus amongst the

face-impressions: beauty, r_M = .68; approach, r_M = .65; like, r_M = .60; danger, r_M = .58. These differences were not significantly different from each other, $F(3, 84) = 1.32$, $p = .27$, $\eta_p^2 = .04$, $\eta^2 = .04$. The next level of consensus occurred amongst the like-impressions of objects, $r_M = .38$; which was still significantly greater than zero, $t(24) = 2.87$, $p < .01$, $d = .57$. Finally, consensus was not observed for beauty-impressions of novel-objects, $t(24) = .93$, $p = .36$, $d = .19$, $r_M = .03$. Multiple comparisons between conditions are reported within the appendix.

The prior results show that impressions are consistent and can be shared, so we should expect the very first impressions to be shared too. We examined impressions made during the first exposure and found the resulting pattern of significance to be the same; see **Figure 15**. Using *t*-tests against zero, there was significant consensus for object-impressions of danger, $t(24) = 26.53$, $p < .001$, $d = 5.31$, $r_M = .83$; approach, $t(24) = 7.45$, $p < .001$, $d = 1.49$, $r_M = .67$; and like, $t(24) = 2.83$, $p < .01$, $d = .57$, $r_M = .26$. There was also significant consensus for every type of face-impression (beauty, $t(18) = 11.39$, $p < .001$, $d = 2.61$, $r_M = .52$; approach, $t(26) = 14.90$, $p < .001$, $d = 2.87$, $r_M = .50$; like, $t(20) = 12.45$, $p < .001$, $d = 2.72$, $r_M = .48$; and danger, $t(20) = 10.00$, $p < .001$, $d = 2.18$, $r_M = .44$). Again, only the object-impressions of beauty were not significantly shared, $t(24) = 1.21$, $p = .24$, $d = .24$, $r_M = .08$.

3.2.3.3 Shapes

With the presence of consensus, it is possible to meaningfully estimate the extent to which shape-parameters contribute to shared impressions of novel objects. Using linear mixed-effects regressions, we regressed the impressions of objects upon their shapes for

the three evaluations, which showed consensus: danger, approach, and likability. Dangerous-evaluations of *family-x* yielded the greatest consensus and, consequently, the shape-parameters accounted for a high proportion of the variance, $R^2_{fixed} = .66$, $R^2_{total} = .86$. As *points* and *flux* increased in value, objects from *family-x* appeared more dangerous, $\beta_{points} = .41$, $p_{points} < .001$; $\beta_{flux} = 1.60$, $p_{flux} < .001$; $\beta_{points:flux} = .29$, $p_{points:flux} < .001$. Similarly, the shape-parameters predicted approach-evaluations, $R^2_{fixed} = .54$, $R^2_{total} = .74$; and like-evaluations, $R^2_{fixed} = .10$, $R^2_{total} = .18$. Thus, in support of our hypothesis, we observed shapes holding a systematic and isomorphic relationship with impressions of novel objects when there was consensus.

3.2.4 Discussion

Our results demonstrate evidence of shapes' contribution to the nascent impressions people have of novel objects. In absence of additional information, it appears that first impressions, driven by the shape of a novel-object, are lasting impressions. This result is congruent with an observation made by McManus (1980), who observed people evaluating rectangular shapes consistently two years later. Another result we observed was that these impressions were sometimes shared across people; i.e., consensus for novel objects is possible. To the extent consensus was present across people, their impressions were indeed isomorphic with the underlying parameters that define the shape of each object. What this means is that people are able to share impressions of novel objects, at least in part, because of their shape.

Across our studies, we did observe a few participants with low levels of consistency. Low consistency could be interpreted to mean that the task of evaluating

novel objects is not meaningfully engaging, or that participants were not paying proper attention in the experiment. Because one of the aims for this study was to assess the extent for which evaluating novel objects constitutes meaningful behavior, no participants were excluded—even if they demonstrated low levels of consistency. By not excluding participants with low levels of consistency, our estimates are therefore conservative. The results of our study suggest that people generally find the task of evaluating novel objects to be meaningfully engaging.

We found that people generally agreed which novel objects were more or less dangerous, but not which objects were more or less beautiful. This pattern is somewhat congruent with prior literature examining abstract novel objects, wherein consensus occurred amongst approach/avoid behavior (Duckworth et al., 2002), and did not occur amongst preferences (Vessel & Rubin, 2010). However, interaction effects of objects and evaluations have been observed before. For example, different rectangular proportions are preferred depending on the type of evaluation solicited (Hekkert et al., 1994; Russell, 2000). It is therefore possible that certain types of shapes have more relevance for certain types of evaluations.

The novel objects we tested were all generated from a single shape-family, *family-x.* The underlying procedural generation algorithm describing *family-x* was designed only with the intention of generating objects not resembling anything in particular—including standard geometric shapes or geons (Biederman, 1987). *A posteriori* we learned that *family-x* bears similitude with the "radial frequency" (RF) algorithm used to investigate the bouba-kiki (formerly maluma-takete; Holland & Wertheimer, 1964) phenomena (Y.-C. Chen et al., 2016). The gist of this phenomenon is

that there exists consensus from some objects generated from RF matching the sound "bouba" and others matching the sound "kiki." In the formulation of RF, three shape-parameters (*amplitude*, *frequency*, *spikiness*) distort a 2D outline of a circle. RF's *frequency* has effects comparable to *family-x's points* — it varies the number of distal extensions. RF's *amplitude and spikiness* together result in effects comparable to *family-x's flux*. Between the two shape-families, there are two major differences: 1) RF features radial symmetry, while *family-x* does not, because its distally extending points are positioned randomly, 2) RF generates 2D line drawings, while *family-x* generates 3D volumes.

The design of the RF algorithm was inspired by prior research (i.e., Gallant et al., 1993, 1996) demonstrating neuronal tuning in V4 to parametrically manipulated polar, hyperbolic, and Cartesian gratings (Wilkinson et al., 1998). Considering the similarity of *family-x* to RF and its precursor (i.e., polar, hyperbolic, and Cartesian gratings), it could be inferred that the reliability of impressions elicited by the objects generated from *family-x* can be explained by the same neural mechanisms. Namely, consensus regarding which objects from *family-x* seems more or less dangerous, approachable, beautiful, or likable could rely on neuronal tuning at V4.

Nevertheless, it is possible that the underlying shape-parameters of *family-x* represent a particular set of shapes that exerts unique effects upon impressions. Since the primary aim of this paper is to document the reliability of a general effect that shapes may have on impressions, it was therefore necessary to examine more types of shapes. By comparing a larger number of shape-families, it is also possible to examine the extent to which impressions of novel objects are shared for only specific evaluative terms (e.g.,

dangerous), and always idiosyncratic for others (e.g., beauty). Study-1 provided initial evidence of shape effects on impressions. In the subsequent study, we employed a greater diversity of shapes to test if consensus forms for other types of shapes and if the level of consensus would vary by shape-family for the same type of evaluation.

3.3 Study-2

3.3.1 Introduction

Across two studies, the reliability of object impressions was tested by examining the consistency and consensus that emerges across participants. The first study examined the reliability of four types of evaluations: approachable, beautiful, dangerous, likable. Participants evaluated one shape-family and one set of same-race portraits 10 times. Consistency remained high under every condition. Consensus varied amongst object impressions, but not amongst face impressions. These results suggest that impression of objects is meaningful even if they may not always be shared across people. Regarding shared meaning, it appears to be tied to the type of evaluation being made of a particular stimulus.

In this study, we sought to explicate shape effects upon nascent impressions by including two additional shape families: *family-y* & *family-z*. We focused on two evaluations, dangerous and beautiful, because they represented the highest and lowest level of consensus observed in study-1. We hypothesized that consensus would form for other types of shapes and that it would result from an interaction between the type of shape and the type of evaluation solicited. Such a finding would further substantiate the

claim that shapes contribute to impressions of novel objects by demonstrating that systematic shape-effects aren't solely specific to certain types of evaluations, nor are they unique to *family-x*.

Additionally, we ran a separate online survey to check the ease, familiarity, and novelty of our task and its stimuli—three presumptions we formerly took in study-1. The ease with which an object can be perceived and associated with semantics relates to its perceptual and conceptual fluency (Reber et al., 2004). Objects which are processed more fluently are evaluated more positively (Reber et al., 2004). Therefore, we checked the baseline fluency of each shape-family using an online survey to determine how the perceived ease or difficulty of an evaluation relates to impressions. We also checked the perceived familiarity of our shape-families, because fluency can also relate to how familiar something seems (Whittlesea, 1993). Finally, we also asked people about the associations which came to their minds in order to check the baseline novelty of each shape-family.

The design of study-2 is similar to study-1, except that we reduced the number of repeated exposures to two, since consistency and consensus emerged in the first two blocks of evaluation in study-1. This alteration allows us to accommodate a larger set of novel-objects without increasing the length of the study.

Study-2 was also designed to confirm the results of study-1 in two ways. First, the initial two blocks of both study-1 and study-2 were the same. Second, we took a portion of the objects from *family-x* and re-rendered them using different settings to test if impressions are consistent despite changes in low-level visual qualities between the different renderings of the same object. Finally, we included a question at the end of the

experiment gauging how likely participants believed others would agree with their own impressions. This final question was included to test whether people's intuitions about the similarity of their impressions to others' impressions are accurate.

3.3.2 Method

3.3.2.1 Participants

Using the same sampling methods as in study-1, 47 volunteers were recruited for study-2. None of the participants have seen any of the objects before. Nine participants were excluded from the analyses because their evaluative ratings lacked variance across entire shape-families. In the final analyses, 20 participants made beauty-evaluations, and 18 participants made danger-evaluations.

3.3.2.2 Stimuli

126 different objects were sampled from *family-x, family-y, and family-z.* **Figure 16** depicts a sampling of objects. As in study-1, each shape-family was designed so as to not appear like anything in particular. *Both family-x* & *family-z* differ on two parameters: *points* and *flux*. The difference between *family-x* and *family-z* is in how the distal points are expressed. *Family-x* produces surfaces which push points distally, while *family-z* produces surfaces which pulls points proximally. *Family-y* incorporates four parameters: *points*, *flux*, *length*, and *radius*. *Point* dictates the number of theoretically distal points. *Flux*, in this case, shifts the points slightly. *Length* dictates the magnitude by which the points are projected relative to its volume. *Radius* dictates the size of a cylindrical edge formed at each distal point.

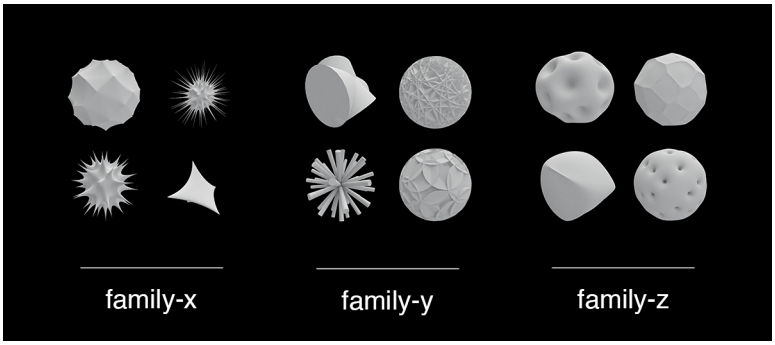


Figure 16. Sample of three shape-families.

Thirty objects were sampled from *family-x, family-y, & family-z.* Each family defines a morphological relationship amongst its objects. *Family-x* is the same shape-family used in study-1.

156 images were used in this study; 96 images of 66 object sampled from *family-x*, and 30 images of 30 objects sampled from both *family-y* and *family-z*. The reason for the greater number of images than objects of *family-x* is because we re-rendered the objects. The purpose of doing that was to check if the impressions were attributable to perceptual properties of the 2D renderings rather than the morphological properties of the 3D objects. **Figure 17** exemplifies the differences between the renderings.

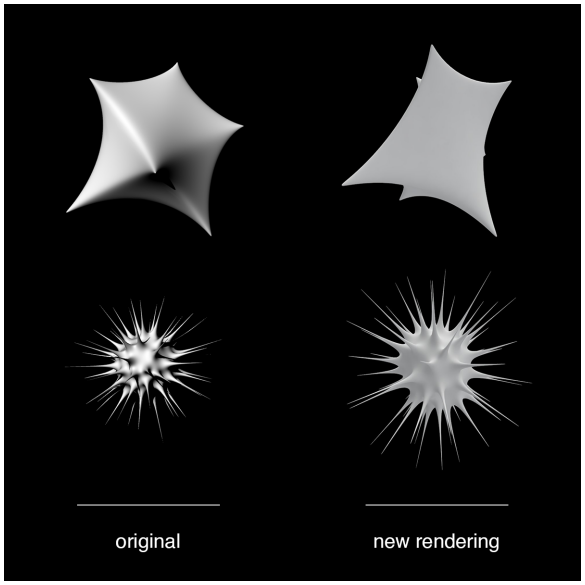


Figure 17. Comparing study renderings. Of the original 66 objects sampled from family-x, 30 objects were rendered as new images. Horizontally, these are the same objects, vertically these are different renders. Despite differences between renderings, impressions of each object were significantly correlated.

The new renderings in this study were created using V-Ray (version 2; www.chaosgroup.com). Rhino's rendering engine, previously used in study-1, doesn't share the specific Bidirectional Reflection Distribution Function (BRDF) of the default rendering engine. BRDF describes the ratio of a surface's radiance, captured by a renderer's camera, to the surface's irradiance brought on by a renderer's light source (Bartell et al., 1981). Due to the lack of technical specification, we moved to using V-Ray. V-Ray also simulates more physically correct illumination of 3D surfaces than the previous renderer used in study-1. The diffuse color of the shader was gray (R = 190, G =190, B = 190). Similar to the previous settings, the refraction and reflection were set as off, and the IOR was 1.0. The BRDF can be specified using V-Ray, and was left using the default settings (type = Blinn, anisotropy = 0, rotation = 0, soften = 0, fix dark glossy edges = on, & derivation = UVW). The rendering camera in V-Ray can also be specified:

F-number = 8.0, ISO = 400; distortion = 0, type = still camera, shutter speed = 200, zoom factor = 1.0, lens shift = 0, vignetting = 0, & exposure = on. The indirect illumination settings were left with their default settings, this included the primary light bounces being calculated using an irradiance map (min rate = -3, max rate = -2, color threshold = 0.4, normal threshold = 0.3, distance threshold = 0.1, hemispheric subdivs = 50, interpolation samples = 20, interpolation frame = 2, interpolation type = least squares fit, sample look up = density-based, calc. pass interpolation samples = 15, multipass = on, randomized samples = on, check sample visibility = off) and the secondary light bounces being calculated using a brute force method (subdivs = 8, bounces = 3). Options for calculating caustics were unnecessary and therefore set as off. The default illumination was used, which includes a skylight that provides indirect illumination for the 3D models; it was set with a multiplying value of 2 to increase its brightness. These new renderings were 1000 x 1000 pixels and presented at a 15° visual angle. The 2D renderings of the objects appeared neutral gray and were presented on a black background.

3.3.2.3 Procedure

A random object from a shape-family was presented for evaluation. Participants were asked either: “how beautiful is this object” or “how dangerous is this object?” Again, each participant engaged in only one type of evaluation during their session. The first two blocks replicated the previous study using the original stimuli. Thereafter, the new renderings were tested. Each of the three shape-families was presented in a counterbalanced order, repeating only once more after each family was seen; see **Figure 18**.

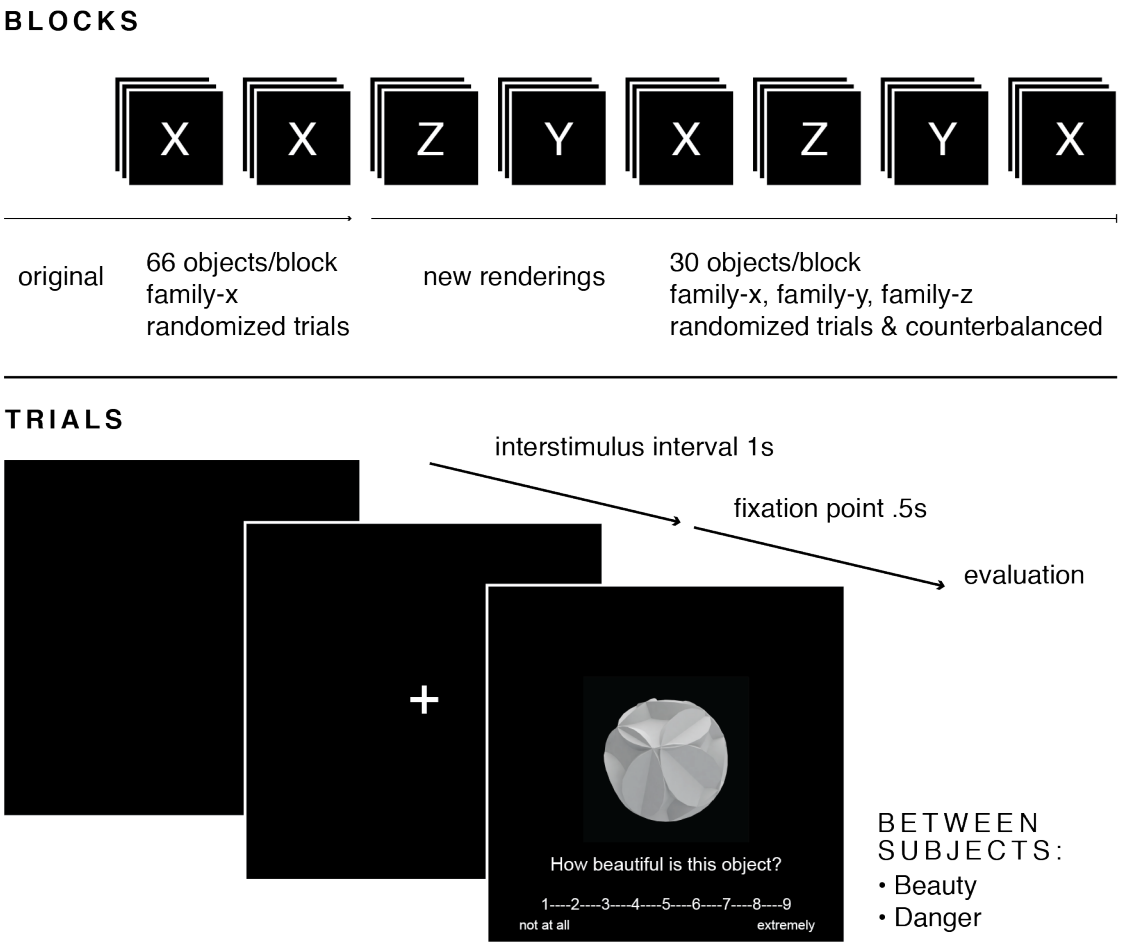


Figure 18. Study-2 Experimental Design. Similar to study-1, random objects were presented for evaluation in a blocked design. These blocks were repeated to allow for measurement of consistency. Each block included objects from one shape-family. The first two blocks replicated study-1. The later blocks featured new renderings of *family-x*, as well as two new *families: y & z.*

At the end of this experiment, participants were asked: "to what extent do you think others would agree with your evaluations?" They responded by keying a number on a scale from 1 (completely disagree) to 5 (neither agree, nor disagree) to 9 (completely agree).

3.3.2.4 Main Analyses

The primary focus of study-2 was to test a more diverse set of objects. We again calculated consistency and consensus statistics based on Pearson's *r* correlation coefficients. We use Fisher's *r* to *z* transformations to calculate standard errors and test for significance. Plots incorporate *r* values transformed back from *z*-scores to display the means with standard error bars, but otherwise the untransformed *r* values are reported. Finally, linear mixed-effects models were again calculated to verify if the shape-parameters systematically predict impressions. For more specific details regarding these calculations, refer to the previous method section of study-1.

3.3.2.5 Manipulation Checks

3.3.2.5.1 Replicating study-1

Study-2's experimental design replicates part of study-1. The original stimuli were presented in the first two blocks. Beauty and dangerous evaluations were examined again. Based on the results of study-1, we expected there to be significant consistency for both evaluation types, however only dangerous-evaluations would result in significant consensus.

3.3.2.5.2 Manipulation Check: 2D images vs. 3D objects

Other variables aside from 3D shapes (e.g., image composition, surface reflectance, or memory for the stimuli) could explain the reliable impressions. To rule out these possibilities, 30 of the 66 objects from *family-x* were re-rendered with a different size, surface quality (e.g., illumination, reflectance, shader, etc.), and at a different angle. If impressions remained consistent despite the visual differences between renderings, participants were likely responding to the morphological 3D properties rather than to

specific qualities found within each of the 2D images. For each participant, we calculated the Pearson's *r* correlation between their average impressions of both renderings. We report the average correlation, representing our measure of consistency across renders, across all the participants.

3.3.2.5.3 Ease, familiarity, and novelty

Three presumptions from study-1 were checked using an online survey: 1) evaluating novel objects seems reasonable and not too difficult; 2) the objects were novel insofar as they did not seem significantly familiar or unfamiliar to people; 3) the objects are novel insofar as different people have different associations which come to mind. Using Amazon Mechanical Turk (mTurk), 60 participants were surveyed online. However, data from only 57 participants were used in the final analyses because three participants saw the stimuli in another online experiment or had incomplete data. The stimuli consisted of the same "new renderings" used in study-2. For each shape-family, every object used in study-2 was presented together in a 5x6 grid on a black background, see **figure 9**. The placement of the objects was randomized across participants.

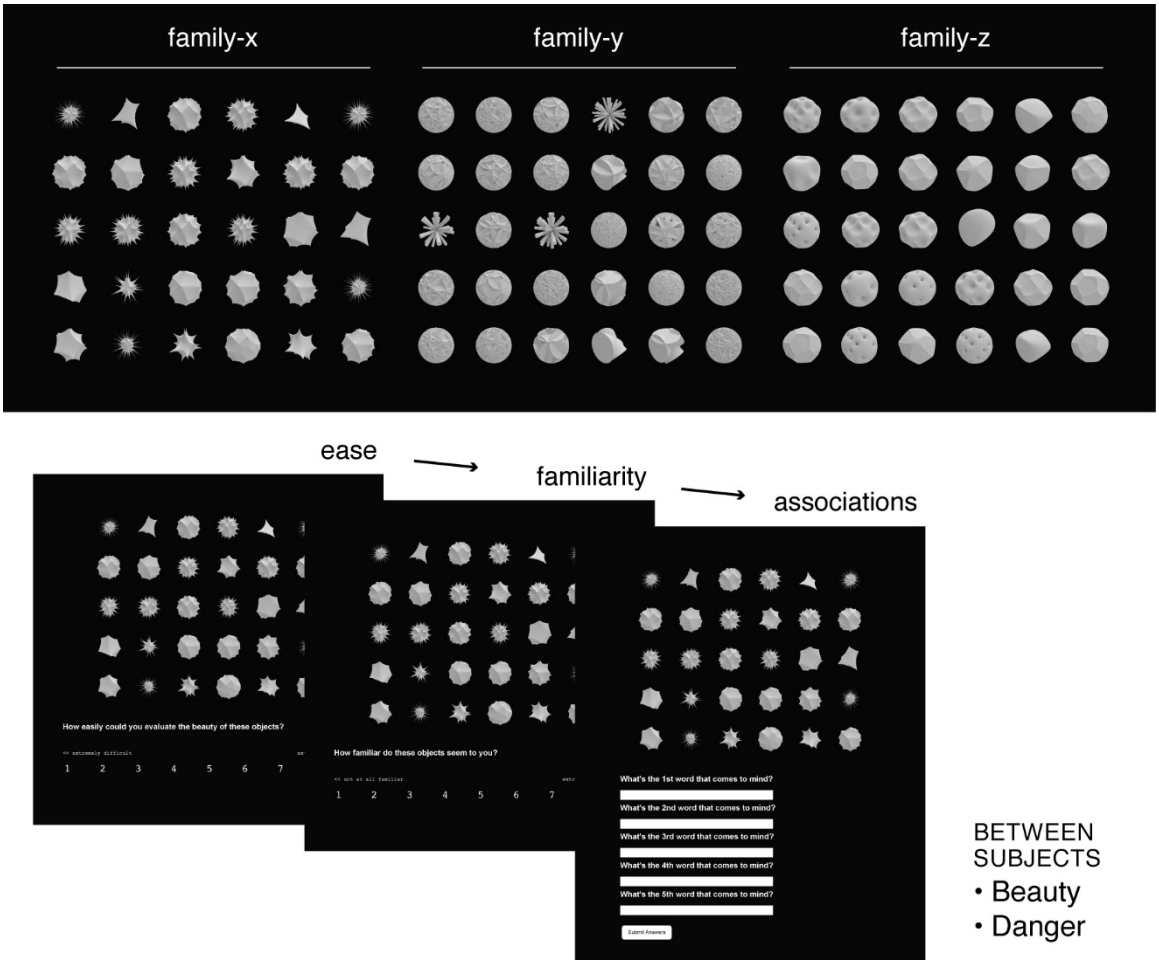


Figure 19. Manipulation Check: Experimental design for checking ease, familiarity, and novelty. Shape-families were presented as a grid. We checked each shape-family for how difficult they were to evaluate, how familiar they seemed, and the diversity of associations they elicit.

There were two evaluation conditions tested between participants. One half of the participants was asked: "how easily could you evaluate the beauty of these objects?" While the other half was asked: "how easily could you evaluate the danger of these objects?" Responses could range from 1 (extremely difficult) through 9 (extremely easily). Then the shape-families were presented again, but this time with the question: "how familiar do these objects seem to you?" Responses could range from 1(not at all familiar) through 9 (extremely familiar). Finally, each shape-family was presented a third

time and participants were asked to list in order the first five associative words which came to mind. Each shape-family remained visible until a response was recorded.

3.3.3 Results

3.3.3.1 Manipulation Checks

3.3.3.1.1 Replicating Study-1

Confirming our findings of study-1, consistency and consensus from beauty-evaluations and danger-evaluations did not meaningfully change; **Table 1** lists the average *r*-correlation values across studies. Consistency was not significantly different across studies for beauty-evaluations, $t(40.28) = .45$, $p = .65$, and for danger-evaluations, $t(32.60) = .24$, $p = .82$. Consensus was not significantly different across studies for beauty-evaluations, $t(33.16) = -1.49$, $p = .15$, but it was significantly lower in the second study for danger-evaluations, $t(38.68) = 4.82$, $p < .001$, $d = 1.49$. However, despite the lower level of consensus in the replication, a *t*-test against zero revealed that consensus for danger-evaluations in study-2 was nevertheless highly significant, $t(17) = 19.87$, $p < .001$, $d = 4.68$, $r_M = .86$.

Table 1. Replication of study-1.

(Study-2) Beauty & Danger evaluations of novel objects.

	N	Consistency (r_M)	Consensus (r_M)
Beauty			
Study-1	25	.58	.03
Study-2	20	.55	.14
Danger			
Study-1	25	.77	.94
Study-2	18	.76	.86

Notes. The average Pearson's r correlation coefficient representing consistency and consensus is reported.

3.3.3.1.2 2D images vs. 3D objects

Despite changes between the two different renderings of the same 3D objects, people formed consistent impressions. Correlating impressions of the old renderings with impressions of the new renderings, the level of consistency was significant for both beauty-evaluations, $t(19) = 7.76$, $p < .001$, $d = 1.74$, $r_M = .55$; and danger-evaluations, $t(17) = 15.10$, $p < .001$, $d = 3.56$, $r_M = .85$. These results suggest that the impressions our participants reported were likely in response to the differences in shapes across the 3D objects, rather than to specific qualities of the 2D images.

3.3.3.1.3 Ease

We first checked the baseline fluency of each condition by asking people to report the ease or difficulty in evaluating a shape-family. We calculated planned comparisons for each shape-family and evaluation-type using *t*-tests against the midpoint. Danger-evaluations of *family-x* were thought to be significantly easy, $t(28) = 5.07$, $p<.001$, $M = 6.55$, $d = .94$. Meanwhile, all other conditions were thought to be neither significantly easy or difficult. The possible implications for the danger-evaluations of *family-x* are explored further in the discussion section of study-2.

3.3.3.1.4 Familiarity

A second check for imbalanced fluency was to measure the perceived familiarity of each-shape family. We ran separate analyses to capture the potential priming effects of mentioning beauty or danger in the prior question posed to the participants. We again calculated planned comparisons using *t*-tests against the midpoint. Participants in the

danger-condition found *family-z* to be significantly familiar, $t(28) = 2.36$, $p = .03$, $M = 5.66$, $d = .44$. However, all other conditions were not significant. Thus, each shape-family was again relatively similar in terms of fluency. This time, the one exception being those evaluating the danger of *family-z*; the possible implications are also explored in the discussion section of study-2.

3.3.3.1.5 Novelty

Finally, we checked if the different shape-families indeed elicit diverse associations. Across our sample of 57 participants, we recorded 406 unique words in association to the three shape-families. *Family-y* elicited the most diverse associations; eliciting 187 different words. This was followed by *family-x*, eliciting 170 words, and then by *family-z*, eliciting 151 different words. These numbers do not add up to 406 because a small subset of words was associated with all three shape-families. Our presumption was confirmed— the objects are novel insofar as they elicit diverse associations across people. In fact, none of the shape-families elicited the same association across all participants, see **table 2** for list of the most frequently associated words.

Table 2. Associations to the novel objects.

(Study-2) Most frequently associated words for each shape-family.

	Family-X		*Family-Y*		*Family-Z*	
	word	%	word	%	word	%
Most frequent association	Sharp	8	Round	5	Round	6
	Pointy	4	Ball	4	Cheese	5
	Spiky	4	Moon	3	Rock	5
	Star	3	Rubber	2	Moon	4
	Spikes	2	Circle	2	Holes	3

Most frequent first association	Sharp	9	Round	9	Cheese	11
	Star	9	Rubber	9	Smooth	7
	Stars	7	Moon	5	Cookie	5
	Pointy	5	Paper	5	Rocks	5
	Spikes	5	Snow	5	Round	5
Most frequent associations primed by beauty, $N = 28$	Sharp	7	Round	7	Round	7
	Star	4	Circular	3	Cheese	5
	Pointy	4	Moon	3	Rock	4
	Spiky	3	Rubber	3	Holes	4
	Abstract	2	Ball	2	Moon	4
Most frequent associations primed by danger, $N = 29$	Sharp	8	Ball	6	Cheese	6
	Pointy	5	Paper	3	Rock	6
	Spiky	4	Round	3	Round	6
	Pain	3	Circle	3	Moon	5
	Spikes	3	Moon	3	Clay	3

Notes. N = 57 Each participant wrote 5 associations; therefore, the maximum possible number of unique words is 285 for each shape-family. Percentages were rounded up to the next integer.

3.3.3.2 Main Analyses

3.3.3.2.1 Consistency

We again observed individuals experiencing their impressions consistently. Collapsing across the different types of evaluations and shape-families, a *t*-test against zero showed that the level of consistency was significant, $t(37) = 21.07$, $p < .001$, $d = 3.42$, $r_M = .80$. The level of consistency did not significantly differ between beauty-evaluations and danger-evaluations, $F(1, 36) = 3.10$, $p = .087$.

3.3.3.2.2 Consensus

Crucially, we observed varying levels of consensus for both evaluations depending upon the shape-family being evaluated. That is, the level of consensus amongst impressions is not solely dictated by the type of evaluation solicited. An ANOVA revealed a main effect of the evaluation-type upon the level of consensus, $F(1, 36) = 25.28$, $p < .001$, $\eta^2 = .16$. However, as predicted, the effect of the evaluation-type was qualified by an interaction between the evaluation-type and shape-family, $F(2, 72) =$

29.42, $p < .001$, $\eta^2 = .37$. There was no main effect of the type of shape-family, $F(2, 72) = 2.56$, $p = .08$, $\eta^2 = .05$.

Given the significant interaction between the type of evaluation and shape-family, we tested each condition for a significant level of consensus; **Figure 20** plots the level of consensus for each experimental cell. Echoing the results of study-1, *family-x*, again led to significant levels of consensus amongst danger, $t(17) = 17.19$, $p < .001$, $d = 4.05$, $r_M = .89$; and yet not amongst beauty, $t(19) = -.21$, $p = .84$, $d = 0.05$, $r_M = -.01$. What is new to study-2 are the two shape-families: *family-y* and *family-z.* Both of these shape-families yielded shared impressions regardless of the type of evaluation. For *family-y*, the level of consensus was significant for both: danger, $t(17) = 5.39$, $p < .001$, $d = 1.27$, $r_M = .53$; and beauty, $t(19) = 6.35$, $p < .001$, $d = 1.4$, $r_M = .60$. For *family-z*, the level of consensus was lower, but nonetheless significant for both: danger, $t(17) = 2.95$, $p = .009$, $d = 0.70$, $r_M = .31$; and beauty, $t(19) = 4.35$, $p < .001$, $d = .97$, $r_M = .42$. These results therefore show that other types of shapes, beyond those represented in *family-x,* can therefore elicit shared

impressions, and that the level of consensus varies across different types of evaluations and shapes.

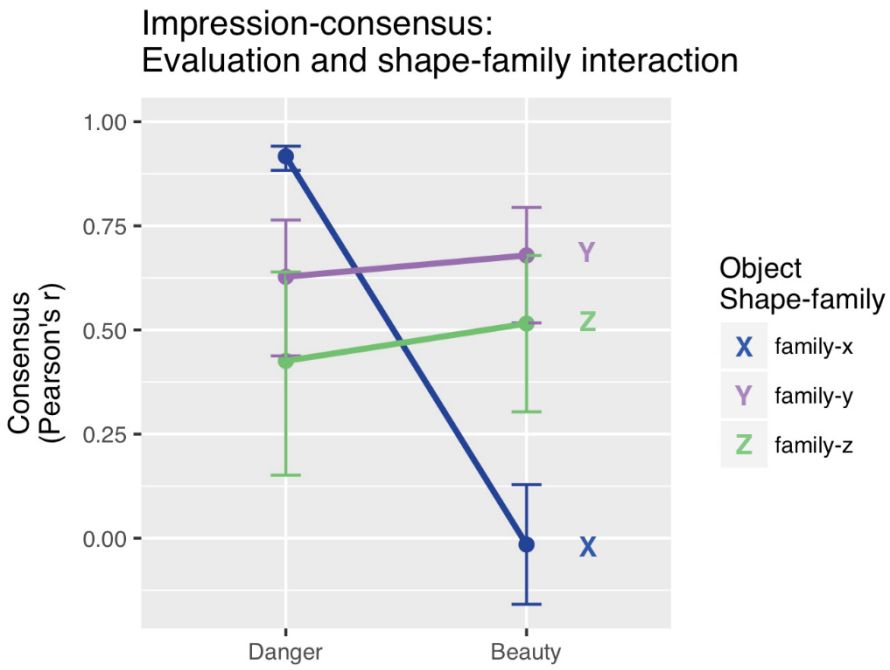


Figure 20. Significant interaction effects when assessing consensus. Consensus varies depending upon on the type of evaluation and shape. The mean consensus values are plotted along with standard error bars representing the .95 confidence limits.

3.3.3.2.3 Shapes

This study included a greater variety of shapes than study-1 to further test whether differences amongst impressions and shapes-parameters are isomorphic to the extent that consensus emerges. Following study-1, we computed a regression model for each type of shape-family and each evaluation which resulted in significantly shared impressions. The greatest consensus occurred amongst danger-impressions of *family-x* and therefore as predicted, the fixed effects (i.e. shape-parameters) explained a large proportion of the variance, $R^2_{fixed} = .53$, $R^2_{total} = .80$. The trend between the level of consensus and the variance explained by the fixed-effects continued with the impressions of beauty from

family-y, R^2_{fixed} = .23, R^2_{total} = .61; danger from *family-y*, R^2_{fixed} = .15, R^2_{total} = .61; danger from *family-z*, R^2_{fixed} = .10, R^2_{total} = .50, and beauty from *family-z,* R^2_{fixed} = .09, R^2_{total} = .68. Thus, the effects shapes have on impressions of novel objects are readily apparent when the impressions are shared.

3.3.3.3 Intuitions about Consensus

Upon finishing the experiment, we asked participants to assess the extent to which others would have similar impressions of novel-objects. On average, participants did believe they would share similar impressions with others, M = 6.16 (on a scale from 1 to 9). There was no significant difference between the different types of evaluations, $F(1, 36) = 0.90$, p = .35. Furthermore, for all conditions (i.e., evaluation-type and shape-family), there was no significant relationship between a participant's belief that their impressions were similar to others and actually having similar impressions to others. We tested this by regressing the extent for which each participant contributed to consensus (i.e., correlation with the mean impressions of other participants) upon their intuition for consensus; see **Figure 21** for a plot of these linear regressions. While evaluating how beautiful or dangerous something is, participants were inclined to believe that others would provide similar evaluations to their own. However, these intuitions were not accurate. That is, although we observed general consensus, participants overestimated the similarity their own impressions would have with those from others.

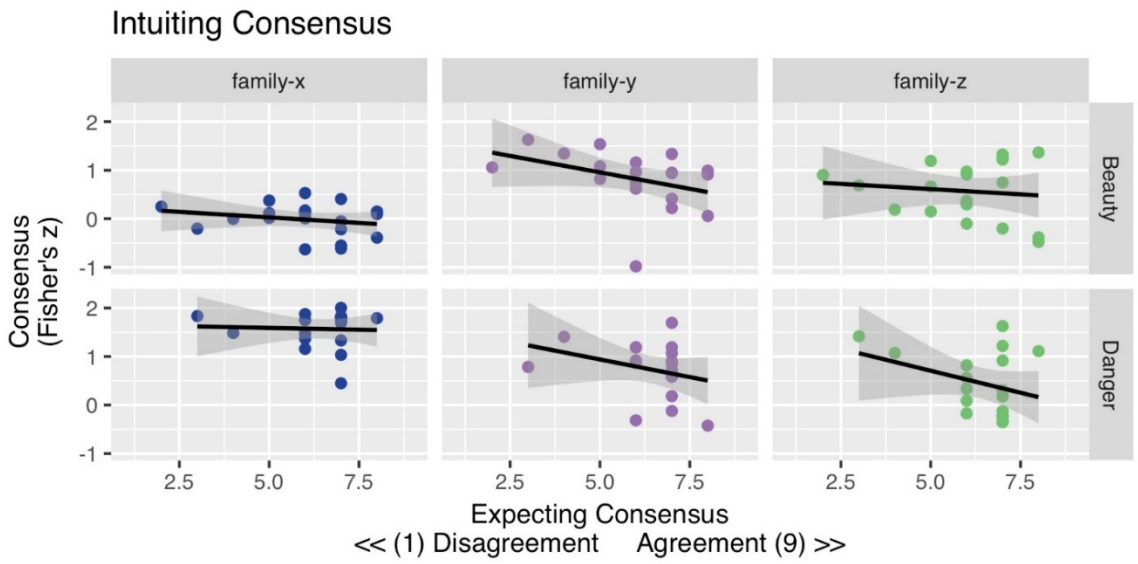


Figure 21. Intuitions about the degree to which others share their impressions are not necessarily accurate. Plotting the Fisher's *r* to *z* transformed scores against people expectation for agreement amongst impressions; there was no significant relationship in any condition.

3.3.4 Discussion

Consensus for novel objects depends upon an interaction between the type of evaluation being solicited and type of shapes being observed. Although in study-1 we observed impressions of danger to be highly similar across people, study-2 showed that the extent to which these impressions are shared depends on the types of shapes being evaluated. Furthermore, the claim that some impressions (e.g., dangerous) are always less subjective than others (e.g., beautiful) is evidently falsifiable. Contrarily, we observed similar levels of consensus amongst beauty-impressions and danger-impressions with some shape-families. A more appropriate claim would be that different impressions vary in consensus depending upon the types of shapes being viewed.

Not only did the average level of consensus vary across conditions, the standard error around the mean was also highly variable. For example, danger-evaluations of

family-x not only results in a higher level of consensus than danger-evaluations of *family-z,* but also results in a smaller range of standard error. Because of the way in which consensus is calculated, the variability in standard error is attributable to participants reporting impressions that are in more or less accordance with the average (i.e., normative) impressions. It is possible that this method of calculating consensus using the "correlation with the mean" inflates the consensus relative to the "pairwise correlation" method. For example, consensus amongst danger-evaluations of *family-z* results in a much weaker correlation using the "pairwise" method, $r = .11$, than "correlation with the mean" method, $r = .31$. We focused on the latter method because it allows for tests of significance. However, regardless of the method, the primary inference made in study-2 remains intact — the levels of consensus are not solely dependent upon the type of evaluation, such as danger, but rather varies across different types of evaluations and shapes.

Amongst impressions of dangerousness, we observed the most consensus with *family-x*, and the least consensus with *family-z*. It is possible that this is related to the different baseline levels of perceptual fluency across these conditions. Recalling the results of our manipulation checks, danger-evaluations of *family-x* was found to be significantly 'easy,' while *family-z*, if primed by the word 'danger,' was found to be significantly 'familiar.' Taken together, this could mean that the ease of evaluating the dangerousness of *family-x* is related to why people are likely to share impressions, while the perceived familiarity of *family-z* is related to why people are not likely to share impressions. Nonetheless, what is important to note here is that there was variation in consensus across beauty-evaluations of different shape-families although they were

balanced in terms of their baseline perceived fluency. Therefore, our results show that perceived fluency alone cannot supplant our hypothesis that shapes contribute to shared impressions.

By testing impressions across multiple renderings of the same object, we further demonstrated the importance of shapes in the evaluation of novel objects. Initial impressions were unperturbed by changes in 2D rendering qualities, e.g., image size, surface-shader, and viewing angle. The permanence of shape's effect upon impressions may be supported by neurons selectively tuned to 3D shapes over 2D qualities (Yamane et al., 2008). It is plausible that shapes take precedence over other perceptual qualities when evaluating novel objects. Infants as young as 4.5 months have been found to individuate objects using their shape, months before their ability to individuate objects based on their patterns or colors (Wilcox, 1999). And it has been shown that infants as young as 6 months are already equipped with a region of their brain selectively activated to changes in shape (Emberson et al., 2017).

At the end of the experiment, we assessed participants' general intuitions about the degree to which others share impressions. Participants on average expected to share similar evaluations with others. However, they overestimated the similarity of their own impressions to the impressions of others. Therefore, the expectation for sharing similar evaluations of novel objects with others constitutes a specific instance of the false-consensus effect, i.e., participants overestimate the degree of similarity between them and others (Marks & Miller, 1987).

3.4 Chapter Discussion

3.4.1 Overview

In study-1, we examined four types of evaluations (approachable, beautiful, dangerous, likable), and two types of stimuli (novel-objects, faces). People proved to have consistent impressions across all types of evaluations and stimuli. Consensus, however, was not equally reliable between both types of stimuli. Faces always resulted in a reliably significant level of consensus, while novel objects resulted in varying levels of consensus; with only some types of impressions (dangerous & approachable) being evidently shared. Regardless, we observed evidence in support of our main hypothesis—that the shape-parameters would be isomorphic with shared impressions when there is consensus. In study-2, we replicated the core findings of study-1 while also testing a more diverse set of novel objects to show that consensus is not solely driven by specific types of shapes, such as *family-x*, or by a specific type of evaluations, such as dangerousness. The results suggest that shared impressions could occur for any novel object, depending upon its shape and the evaluation being solicited.

The question of why consensus is only found amongst specific combinations of shapes and evaluations remains an open question. Based on prior literature, there are many possible explanations. Here we offer three in no particular order: 1) Certain shapes have certain "affordances" and consensus occurs when they assess a relevant affordance (for an overview: Chemero, 2003; Jones, 2003). 2) Some shape-families, despite being abstract, bring to mind commonly shared semantic associations and therefore certain types of evaluations will lead to consensus (e.g., Vessel & Rubin, 2010). However, in

study-2 we showed that people have different associations that come to mind for the present set of objects. Although this could serve as contradictory evidence, semantics could still offer an explanation if the diversity of associative words we recorded actually solicits the same semantic network. 3) Based on the notion of "micro-valence" (generally defined as a "subtle affective value" by Lebrecht et al., 2012), each shape-parameter and evaluative-term (e.g., dangerous, beautiful) may be, to varying degrees, positive or negative (i.e., micro-valence). From this perspective, consensus may be explained by the likelihood of a similar mapping of valence of the evaluative-term and shape-parameters. This explanation, along with the others, may be tested in future research by leveraging procedural generation. For example, shape-families could be designed to manipulate affordances, semantics, and valence. Studying such explanations could lead to an understanding of how all shapes may relate to one another, and how impressions of novel objects may be codified.

The results of our studies follow what has long been evidenced in more applied realms of psychological research. While examining which shapes should symbolize aircrafts in the military, rounded figures were more likely to be perceived, easily learned, and accurately recalled as friendly entities (Provins et al., 1957). Specific interactions between fonts and English-words result in faster identification of meaning (Lewis & Walker, 1989). Trees and schematics similar in shape are evaluated similarly (Summit & Sommer, 1999). Curvilinear, over rectilinear, architecture is more likely to be evaluated as being beautiful (Vartanian et al., 2013). Products advertised with a curvilinear, rather than a rectangular, logo are expected to be more comfortable (Jiang et al., 2016). Our

studies contribute to this literature by demonstrating the reliability of shape's effects through the manipulation of the most complex, yet controlled, novel objects to date.

3.4.2 Exploratory PCA

What does it mean when consensus is not found? We reported above that there was no general consensus regarding the beauty of *family-x*. Using exploratory principal components analysis (PCA) with varimax rotation on the level of participants, we were able to test the extent to which beauty-impressions were idiosyncratic. If completely idiosyncratic, then there should be an equivalent number of principal components as there were participants entered into the analysis. The results showed four components reliably emerging with an eigenvalue over 1. These four factors explained 87% of the total variance in study-1, and 78% in study-2. Applying a *k*-means clustering analysis, we found a high level of consensus amongst two clusters of participants within both study-1 and study-2. In other words, even if general consensus was not found, impressions were not completely idiosyncratic. Instead, there exists sub-groups sharing similar evaluations. Of the total variance explained by these four-factor PCA models with two clusters, 54.7% of the variance can be explained by the shape-parameters in study-1, and 49.7% in study-2. It remains unknown what causes people to evaluate objects in more or less similar ways. The results of this exploratory PCA demonstrate the possibility for studying mediating factors (e.g., social, contextual, situational, personality, prior experiences) whenever general consensus is not found.

3.4.3 Individualized Models

Another way to examine shape's effect when consensus is absent, is to analyze the impressions at the level of every individual. We therefore calculated individualized linear models—every participant's average impression was regressed on the scaled values of the shape-parameters. From each model, we extracted the beta coefficients to plot them against the level of consistency exhibited by the same individual represented by the model. The resulting plots confirm the relevance of shape-parameters even in the absence of consensus, see **Figure 22**.

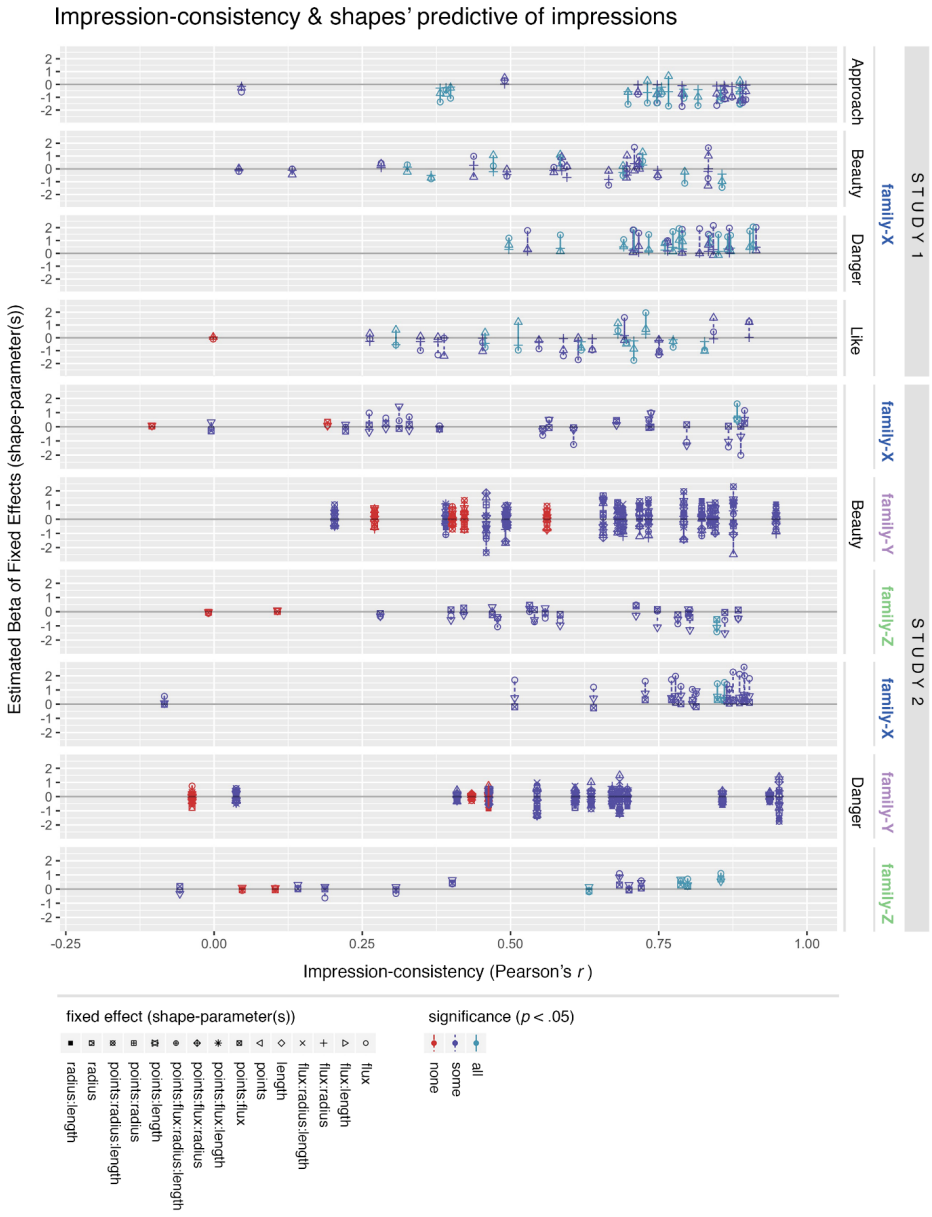


Figure 22. Consistency & shapes' predictive of impressions. Every individual's impressions were regressed on the shape-parameters. Plotted are the beta coefficients (y-axis) against an individual's consistency (x-axis). The beta coefficients describe how participants are individually affected by each shape-parameter, and their interactions, when forming their impression. If at least one shape-parameter was significantly predictive ($p < .05$) of an individual's impressions, it is marked as 'some'; otherwise, the color of a data-point denotes that 'all' or 'none' of the shape-parameters were significantly predictive. Note that in study-2, five data-points could not be plotted because of a lack of variance across impressions of *family-z; one beauty-evaluation, four danger-evaluations*. Using this plot, it is possible to glean a qualitative account of how similar individuals are evaluating the novel objects. Similar patterns of fixed effects mean a similar model of impression. Individual models in which none of the shapes were significantly predictive of impressions coincided with lower levels of consistency.

A relationship can be observed in **Figure 22** between an individual's consistency and the extent for which shape-parameters were significantly predictive of their impressions. Individualized regression models where none of the shape-parameters were significantly predictive of impressions tended to be models of impressions with poor consistency. A prime example occurred in study-1, where one participant's impressions of likability of *family-x* was not predicted by the shape-parameters and also was not consistent (r_M = .00) across re-evaluations.

Another relationship observable in **Figure 22** is the pattern for the shape-parameters. For example, the pattern in which impressions map to shape-parameters appears regular across participants for danger-evaluations of *family-x*, but not beauty-evaluations of *family-y*. Recalling the results of study-2, these two conditions also represented the highest and lowest levels of consensus. Taken together, **Figure 22** provides a qualitative look at how shape-parameters work when consensus is present and when consensus is absent. Amongst beauty-evaluations of *family-y*, *points* positively mapped to beauty for some people, but negatively mapped to beauty for others. These results are congruent with the prior exploratory PCA results in that there appears to be clusters of participants sharing similar mappings between shape-parameters and impression. Thus, even if a participant does not share the same impression of a novel object with most other participants, they may still share their impression with some others because of the shape of the objects.

3.4.4 Summary

How do people form impressions of objects? We hypothesized that shapes contribute to initial impressions. To test this, we first examined the reliability of novel object impressions. Our results show that impressions of novel objects are indeed consistent, and possibly consensual. In other words, people do not need to know what they are looking at in order to feel a certain way about it—and that feeling may be shared. We then demonstrate that shapes are contributing to reliable impressions by showing that shape-parameters predict impressions.

4.

Shared impressions of shapes. [8]

4.1 Introduction

Abstract and novel shapes can lead people to form a similar impression of a novel object (A. Kurosu & Todorov, 2017). What remains unknown are the types of impressions that are likely shared. The purpose of this chapter is to explore the question of how people form shared novel object impressions by building a model of them in what we call *shapeSpace*.

The utility of *shapeSpace* is both applied and basic. In the world of product development, the notion of 'Kansei engineering' (i.e., affective engineering) focuses on processes capitalizing on the emotional aspects of a product's experience (Nagamachi, 1995). Note that this is similar to the notion of 'product semantics' in industrial design (Krippendorff & Butter, 1984). By researching impressions of novel objects in a robust

[8] Work represents an unplublished manuscript by Kurosu, A., Nikolas, G., & Todorov, A.

psychological fashion, the results of our efforts carry the applied potential for ameliorating design tools (e.g., Mothersill & Bove, 2015; Yumer et al., 2015), improving design education (e.g., Pourtalebi & Bagherzadeh, 2012), and the efficacy of communication between designers and laypeople (e.g., So, 2019).

Prior research has explored mappings between specific categories of visual stimuli and their impressions. Most well-known among psychologists is perhaps a model of faces (Oosterhof & Todorov, 2008). Other models, constructed through a process measuring semantic differentials (i.e., Osgood et al., 1957), have led to many specific maps (i.e., latent space) of impressions: interior architecture (Coburn et al., 2020), micro-electronics (Chuang & Ma, 2001), and telephones (Hsu et al., 2000; Yang & Chang, 2012). There have been a few attempts to construct more general models, however they do so by including objects from multiple known categories rather than by studying impressions of novel objects (e.g., Hsiao et al., 2006; Mastandrea et al., 1992). A model of novel object impressions therefore has never been derived and empirically tested.

From a basic research perspective, modeling *shapeSpace* carries potential for elucidating a domain general mechanism for extrapolating meaning from visual stimuli. For example, consider the following research on fluency and complexity (e.g., Reber et al., 2004; Van Geert & Wagemans, 2020), anthropomorphism and animacy (e.g., Epley et al., 2007; Scholl & Tremoulet, 2000), dangerousness (e.g., Bar & Neta, 2006; Connolly et al., 2016), intuitive physics (e.g., Kubricht et al., 2017), size (Konkle & Oliva, 2011), and perceptual completeness (Johnson et al., 2008). Across these topics within psychology, researchers have studied a comparable situation where within a person

evaluates a visual feature. What if these different effects are the result of a common latent space representing impressions of shapes? Thus, we assert that the basic utility of *shapeSpace* is its potential for bridging existing psychological research.

Our approach is data-driven, and our stimuli consists of state-of-the-art procedurally generated shape-families (i.e., highly defined object sets) (e.g., Kurosu & Todorov, 2017). Documentation of our observations are presented as a series of three studies. In study-1, we start with a broad exploration of the potential meaning elicitable by novel visual stimuli. Following the emergent patterns, in study-2 we formulate a model of *shapeSpace*. Then in study-3, we probe how well *shapeSpace* replicates with impressions of objects from different shape-families, impressions of objects within shape-families, and with every person.

4.2 Study-1

4.2.1 Introduction

In study-1, we start by collecting free-response data from a free-association task. Similar in method to that of Oosterhof and Todorov (2008), the intent was to not impose constraints over the responses to the visual stimuli. Our approach contrasts with those approaches, which compile features derived from prior research (e.g., Aaker, 1997), or from interviews with germane experts (Mastandrea et al., 1992). However, as a result, the resulting data are highly diverse. Each word generated by a person may not necessarily represent an association shared by another person. Therefore, the data were processed before features were extracted.

The end goal of study-1 was to construct a list of potential features representing the landscape of potentially shared impressions of novel objects. The free response data were therefore filtered for adjectives used by at least two different people. In a balance between parsimony and representativeness, we used natural language processors (NLP) to cluster the adjectives according to their common semantic associations. Then we extracted features based on clusters of distinctive meaning.

4.2.2 Method

4.2.2.1 Participants

100 volunteers were recruited online using Amazon's Mechanical Turk (Mturk). All volunteers were briefed before providing consent and then debriefed about the purpose of the study after completing the study. Financial compensation was provided for their time. Participation requirements included: adult age, normal vision, fluency in English, and no prior exposure to our stimuli. Two subjects were removed for having possibly seen the stimuli in another experiment, leaving a total of 98 participants (mean age = 33.14, SD = 8.97; 50 female, 48 male) in the analyses.

4.2.2.2 Stimuli

Participants provided responses to images of 20 novel objects belonging to different shape-families. The objects were abstract, rather than symbolic of any known category. To achieve this, we created the objects through procedural generation. Compared to conventional methods of 3D modeling, where 3D geometry is directly manipulated in 3D space, procedural generation is a method of composing scripts (i.e.,

algorithms) generating sets of familial objects called shape-families (Krispel et al., 2015). Objects belonging to the same shape-family appear more similar than objects belonging to different shape-families. By testing objects belonging to different shape-families our method increases the inherent diversity of shapes represented by our stimulus set.

The shape-families were constructed using Grasshopper (version .9; www.grasshopper3d.com). Images of the objects were rendered using V-Ray (version 2; www.chaosgroup.com). The settings of the renderer were identical to what has been used in prior related research (see study-2 in Kurosu & Todorov, 2017); the exceptions being the background color and the size of the renderings. The objects in this study were presented on a white background and the renderings were 300 by 300 pixels; see **Figure 23.**

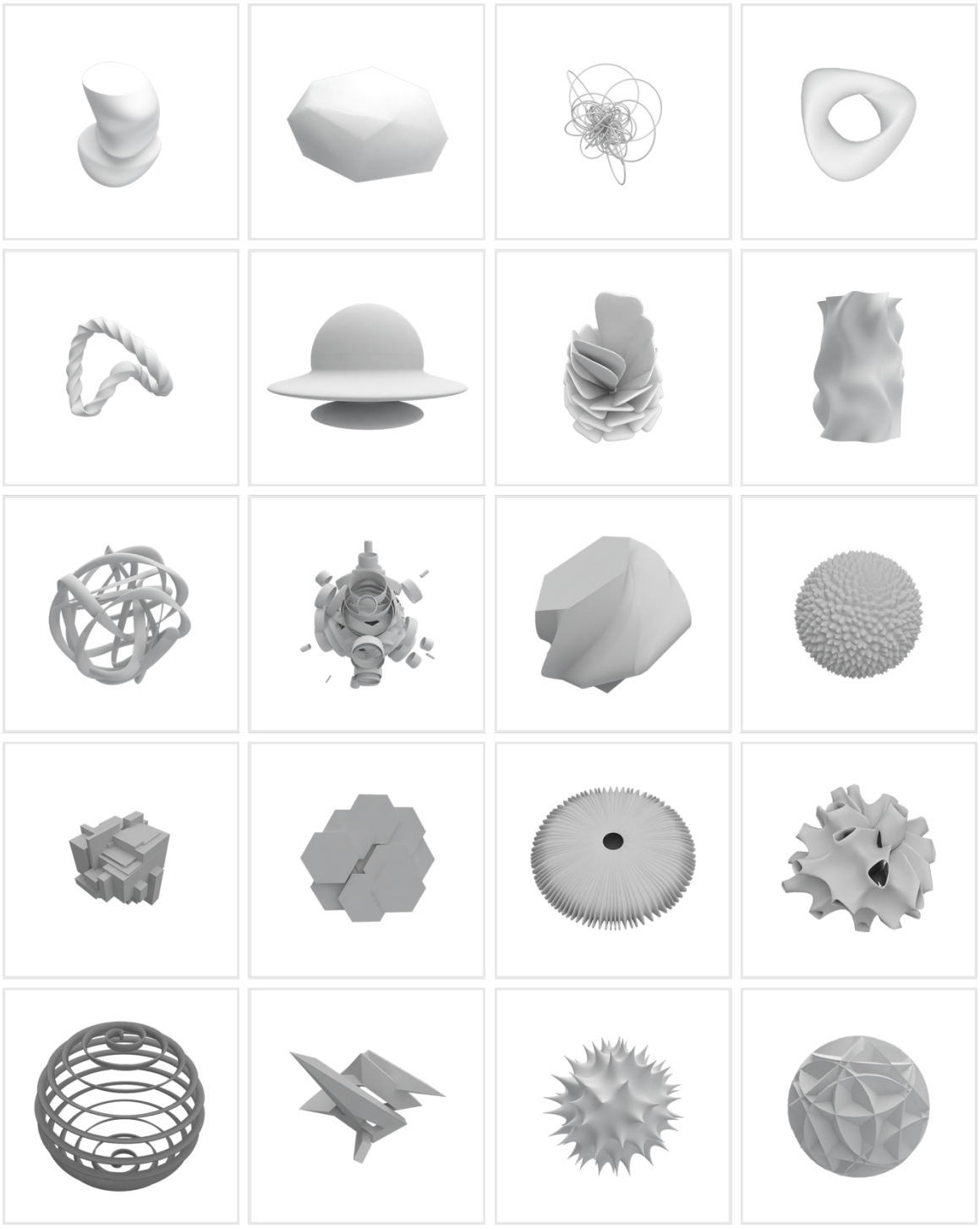


Figure 23. Rendering of each object shown in study-1. Each object belongs to a different shape-family.

4.2.2.3 Procedure

Participants were asked to provide five associations they had to each of the 20 objects. To encourage participants to feel genuinely free to type any association they had, we provided the following broad questions of what they might consider: what does it look like? How does it make you feel? Does it convey an emotion? The instructions emphasized that there was no correct or incorrect answer.

Each object was shown one at a time in a random order. A single object was presented along with a list of five empty boxes. Participants had to provide at least three associations for each object before they could continue to the next trial. Once they moved onto a new trial, they could not go back to change their answer. There were no time constraints for providing their associations. After participants completed the free association task, demographics was collected.

4.2.2.4 Data Filtering

The free association data were filtered for adjectives shared by at least two different people in response to any object. All responses were preprocessed to create more uniformity by removing differences in capitalization, punctuation, and word endings. All analyses and data processing were performed using R (version 3.6.0). Preprocessing was performed using the 'SADCAT' package (version 0.9). Adjectives were identified using the 'koRpus' package (version 0.11).

4.2.2.5 Natural Language Processing

Identifying features (i.e., semantic differentials) relied on finding patterns in the data. We used word embedding and hierarchal clustering to organize the adjectives based on their semantic distances. For a simplified explanation of the intuition: word embeddings models obtain semantic relatedness from large corpora of text based on word co-occurrences (i.e., how often two words appear close to each other) by comparing the similarity of the context in which two words appear. That is, words that tend to co-occur with the same set are considered to be more semantically similar to each other.

For example, both triangle and square often co-occur with words such as shape or sides and are thus highly semantically related; triangle and phone do not often co-occur and therefore receive a lower semantic relatedness score. For every word in, the NLPs used here provided us with ~300-dimensional word embedding vectors. Given that the vectors encode the semantics of words in an Euclidean space, we can obtain a similarity matrix of all pairwise similarities. Words which appear in similar texts are located more closely in the multidimensional vector space. Each word therefore has a location within the space and distances between words can be calculated. The shorter the distance between two words, the more similar they are likely to be.

Three word embeddings models were used: pretrained fastText pretrained on the Common Crawl (https://fasttext.cc/docs/en/crawl-vectors.html; Grave et al., 2018), gloVe pretrained on the common crawl 840B (https://nlp.stanford.edu/projects/glove/; Pennington et al., 2014), and word2vec pretrained on Google News (Mikolov et al., 2013). With each processor, we derived the semantic distance between pairwise

combinations of the adjectives and averaged their distances to decrease the potential bias of using any one particular processor. The output is a matrix representing the similarity (i.e., average semantic distance) between adjectives.

4.2.2.6 Feature Extraction

Features (i.e., semantic differentials) are parameters anchored at each end by two adjectives with opposite meaning (Osgood et al., 1957). For example, positive and negative may represent two endpoints of one feature. Using the similarity matrix described above, the adjectives were organized into hierarchal clusters using the 'hclust' function in R.

The output from the hierarchical cluster analysis positions words so that neighboring words represent relatively more similar meaning. Labels were assigned to clusters of words at various levels within the hierarchy. The labels functions as notes for identifying distinct clusters of meaning that should be reflected in the final set of features, while attempting to reduce the final number of features. Note that our cluster analysis does not take into account frequency, thus our method represents a breath of diverse, yet shared meaning, rather than the most commonly used words.

Amongst some clusters, the constituent words individually were too specific to adequately reflect the meaning shared within the cluster. Therefore, to assemble a feature capturing the general sentiment of a cluster, a new adjective was used. For example, 'Negative v. Positive' is a feature representing a cluster of adjectives that includes annoyed, angry, upset, scared, nervous, nice, good, perfect, neat, cool, and awesome. In the following results section, we provide additional notes documenting the relationship

between free association data and the adjectives represented in the resulting set of features.

4.2.3 Results

The free association task elicited 2730 unique words, 270 of which were shared adjectives; see **Figure 24**. This result demonstrates the diverse ways novel visual information begins to yield meaning.

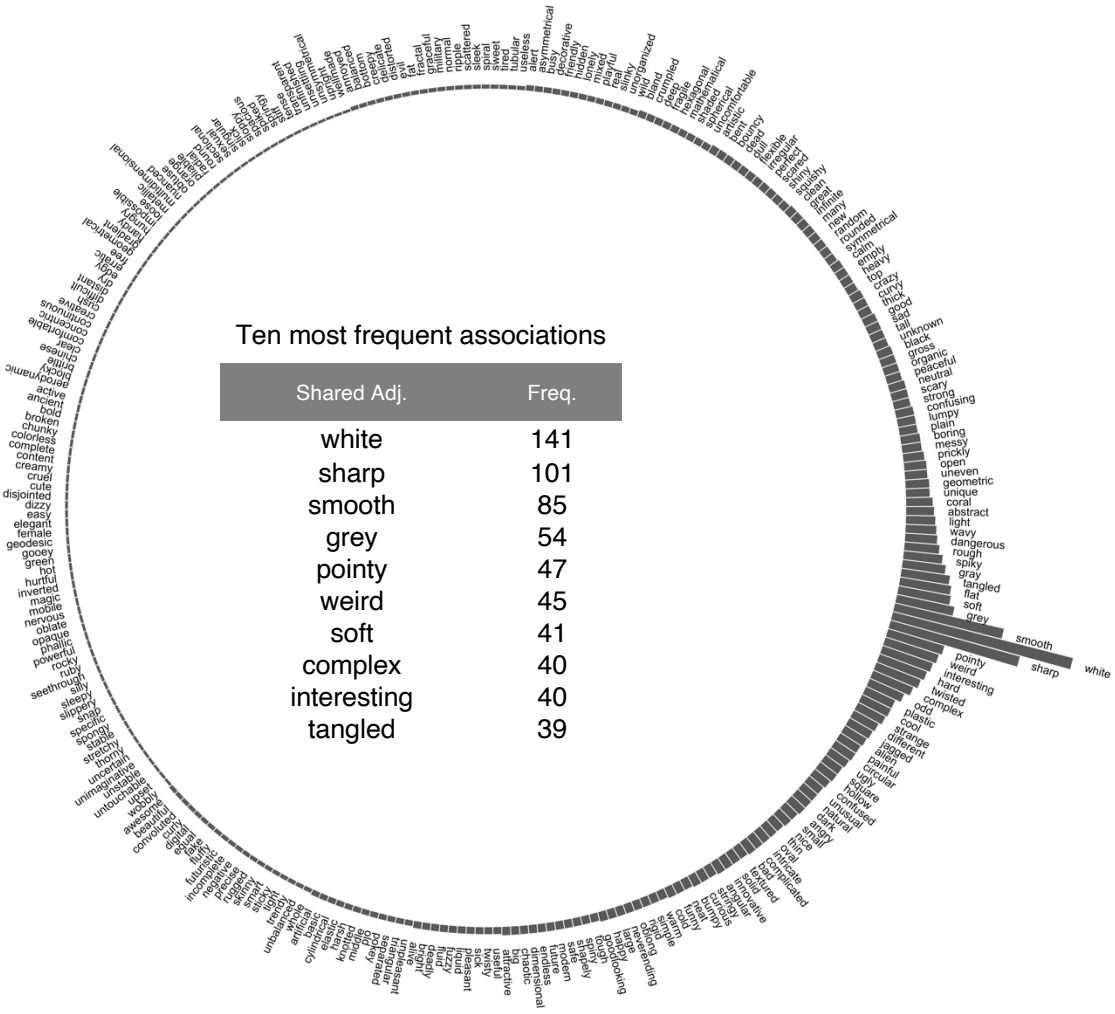


Figure 24. Around the circle is a list of shared adjectives from the free association task. The length of the bars represents their frequency. At the center is a table listing the 10 most frequent associations.

Based on patterns in the free association data, we extracted 57 features; see **Table 3.** Because each feature is comprised of two adjectives, a total of 114 adjectives are represented across the features.

Table 3. Notes regarding extracted features. Some endpoints directly reflect free-association data, while others indirectly represent a cluster of adjectives found in the free association data.

Study-1. Listing features and their relationship to the free association data.

Feature	Direct	Indirect	Freq.
Abstract v. Concrete	abstract	artistic, geometric, dimensional, mathematical	56
Angular v. Curvy	curvy, angular	square, rounded, wavy, curly, squishy, spongy, springy, bouncy, triangular, hexagonal, inverted, geodesic, spherical, cylindrical, tubular, concentric, circular, radial, oval, oblong	150
Asymmetrical v. Symmetrical	asymmetrical, symmetrical	unsymmetrical	11
Calming v. Scary	scary	calm, creepy, unsettling	25
Clear v. Fuzzy	clear, fuzzy	fluffy, pokey	14
Cold v. Hot	cold, hot	warm, dry	19
Dead v. Alive	dead, alive	wild, lonely, distant	21
Discontented v. Content	content	-	2
Discontinuous v. Continuous	continuous	-	2
Disgusting v. Cute	cute	untouchable, gross, sweet	18
Disorganized v. Organized	-	disjointed, chaotic, erratic, unorganized, sloppy, messy	32
Distorted v. Shapely	distorted, shapely	crumpled, twisted	47
Empty v. Full	empty	hollow, plastic	45
Feminine v. Masculine	-	female, sexual	4
Finite v. Endless	endless	neverending, infinite	19
Flat v. Spiky	flat, spiky	shaded, sharp, jagged, pointy, spiked, phallic, thorny, prickly, spiny	261
Incomplete v. Complete	incomplete, complete	unfinished	7
Inelastic v. Elastic	elastic	stretchy, twisty	12
Informal v. Formal	-	playful, bold, edgy, silly, crazy, funny	27
Irregular v. Regular	irregular	wobbly, uneven, lumpy	37
Light v. Heavy	light, heavy	-	29
Liquid v. Solid	liquid, solid	fluid	20
Loose v. Tight	loose, tight	-	5
Mild v. Intense	-	middle, top, bottom, deep, snap	21
Moderate v. Extreme	-	unimaginative, bland, dull, boring, modern, futuristic, unique, innovative, new, future, old, specific, different, many, military	108
Natural v. Artificial	natural, artificial	organic, digital	34
Negative v. Positive	-	annoyed, angry, upset, scared, nervous, nice, good, perfect, neat, cool, awesome	99
Non-slippery v. Slippery	slippery	slick	4
Non-sticky v. Sticky	sticky	gooey, creamy, stringy	16
Partitioned v. Whole	whole	ripple, gradient, sectional	10
Plain v. Busy	plain	shaded	18
Rough v. Smooth	smooth, rough	soft, textured, bumpy, rocky, aerodynamic	174
Sad v. Happy	sad, happy	-	17

Safe v. Dangerous	safe, dangerous	harsh, cruel, deadly	41
Scattered v. Concentrated	scattered	separated, mixed	11
Short v. Tall	tall	upright	12
Sick v. Healthy	sick	tired, hungry, dizzy	12
Simple v. Complex	complex, simple	convoluted, complicated, confusing, confused, obtuse, basic, easy, impossible, difficult, tough, hard	144
Singular v. Multiple	singular	equal	5
Small v. Large	small, large	big	25
Static v. Mobile	mobile	-	2
Stiff v. Flexible	stiff, flexible	rigid, rugged	19
Straight v. Bent	bent	-	6
Thin v. Thick	thin, thick	skinny, chunky, slinky, fat	32
Ugly v. Beautiful	ugly, beautiful	bad, sleek, elegant, shiny, attractive, spacious	50
Unapproachable v. Approachable	-	friendly, comfortable, clean, peaceful	19
Unbalanced v. Balanced	unbalanced, balanced	neutral	23
Uncolored v. Colored	-	transparent, opaque, colorless, ruby, coral, bright, dark, orange, green, grey, gray, white, black	289
Unconstrained v. Constrained	-	neutral	18
Unfamiliar v. Familiar	-	magic, hidden, ancient, evil, alien, fake, chinese, unknown, uncertain	46
Unintelligent v. Intelligent	-	smart	3
Uninteresting v. Interesting	interesting	curious, unusual, weird, strange, odd, random	167
Unpleasant v. Pleasant	unpleasant, pleasant	uncomfortable, painful, tense	33
Unstable v. Stable	stable	active, normal	7
Untangled v. Tangled	tangled	knotted	43
Useless v. Useful	useless, useful	-	8
Weak v. Strong	strong	fragile, brittle, intricate, delicate, powerful	33

Note. A hierarchical cluster analysis based on semantic similarity determined the clustering of adjectives.

4.2.4 Discussion

The results demonstrate a list of features that may be relevant in the evaluation of novel objects. The free associations were first narrowed down to shared adjectives. Then they were organized according to similarity and 57 features were extracted. The approach used here is similar to recent studies exploring stereotypes in a more comprehensive way by leveraging word embeddings to analyze free responses in a more data-driven way (e.g., Nicolas, Bai, & Fiske, under review; in prep.; Nicolas et al., under review).

It should be noted that in using word embeddings, as with all machine learning techniques, the output is not necessarily objective. NLPs are biased by the scientists who constructed them, and the data used to train them. To mitigate this issue, we averaged the

output of three different NLPs. An alternative approach is to recruit participants to judge the relationships between the associations (e.g., Aaker, 1997; Oosterhof & Todorov, 2008). However, such an approach with a list of adjectives numbering in the hundreds would result in a highly impractical and very tedious task. Therefore, the primary reason for using word embeddings is its ability to efficiently identify patterns across large quantities of text.

The primary goal of this study was to identify a list of features representing shared impressions of novel objects. This list is the input to constructing a model of shared impressions using a single multidimensional space.

4.3 Study-2

4.3.1 Introduction

In study-2, we test the reliability of each of the 57 features identified in the prior study. Two measures are used to determine the reliability of judgments of each feature: consistency (i.e., intra-rater agreement, test-retest) and consensus (i.e., intra-rater agreement, shared).

We show participants novel objects and ask them to provide their impression as it relates to one feature. After they rate all the objects, they are asked to do it again and a measure of how consistent their impressions are between both times is taken. Prior research suggests that novel object impressions are significantly consistent, regardless of the feature they relate to (A. Kurosu & Todorov, 2017).

Our second measure of reliability, consensus, is a measure of the similarity in impressions between one person compared with everyone else. Prior research shows that consensus among novel object impressions varies according to the interaction between features and shape-families (A. Kurosu & Todorov, 2017). Alternatively, based on the results of Vessel and Rubin (2010), consensus[9] may be low since our stimuli are essentially a collection of abstract images. Either way, the primary purpose of measuring consensus remains the same. The purpose is to isolate features with stable consensus and model them to build *shapeSpace* and further understand how people share impressions of novel objects.

The objects in this study are digitally interactive. Participants could manipulate them on their computer screen; rotating and zooming to develop a better understanding of their shape. The ultimate goal of *shapeSpace* is to draw inferences about impression formation in a way that generalizes to real world objects. Affording participants some manipulation of the objects was an effort made to minimize the potential ambiguity about shapes as they might exist in reality.

To model *shapeSpace*, we project the reliable features in a multidimensional space using exploratory principal components analysis (PCA). While the brain may represent information in a high-dimensional space (Hasson et al., 2020), we used PCA to compress the features into a lower dimensional space with interpretable (i.e., labeled)

[9] Their study measures agreement uses pairwise correlations, unlike our study of consensus which measures correlation to the mean. Agreement regarding the beauty of abstract images was shown to be $r = .2$, which is lower when compared to faces, $r = .4$.

dimensions to afford the aforementioned basic and applied utility of *shapeSpace*. After modeling *shapeSpace*, we identify a subset of features to focus on while exploring the model further—carrying into study-3. Thus, the goals of the present study are to develop a model *shapeSpace* and identify a subset of features that can replicate its structure.

4.3.2 Method

4.3.2.1 Participants

Data from 1716 participants were used in the following analyses. Volunteers were prescreened for those who were 18 years or older, had normal vision, were fluent in English, had no prior exposure to our stimuli, had access to a modern browser, and a broadband internet connection. The mean age was 34.98 (*SD* = 11.63). 876 were female (825 male, 15 non-binary). 682 completed a bachelor's degree (521 high school, 274 associate's, 195 master's, 20 doctorates', 24 other). 1172 were white (132 Asian, 187 Black, 93 Hispanic, 34 Other, 98 Multiracial). The average participant was artistically naïve, *M* = 8.68, *SD* = 8.90; this was determined using an art experience questionnaire (Chatterjee et al., 2010).

Participants were recruited on Mturk and were financially compensated for their time. We originally recruited 1729 volunteers for this study, however 13 were excluded because they did not have normal vision (despite the pre-screen), experienced a technical problem during the study (e.g., objects not loading, computer crashing), or because their data lacked variance (*SD* = 0). Participants were randomly assigned to one of the 57 features. With regards to the 1716 participants included in our data, on average, each feature had a sample size of 30, *SD* = .67, *min* = 29, *max* =32.

4.3.2.2 Stimuli

Stimuli in the present study were similar to the stimuli in study-1. However, instead of presenting the images as static, they were presented as interactive digital 3D objects. By using a keyboard and mouse, participants could rotate and zoom each object. To implement this feature, we used state-of-the-art WebGL technology by implementing p5 (version 0.8.0), a JavaScript library which facilitates real-time rendering of 3D objects with a web browser.

In order to present the objects using p5, the set of objects used in study-1 had to be modified. Some of the shape-families had to be rebuilt because of the way p5 handled the rendering of *normal* geometry. While rebuilding some of the shape-families, the software we were using, Grasshopper, moved out beta (becoming version 1). This necessitated other aspect of the algorithms be rebuilt too. The final stimuli used in the present study bore a great degree of similitude to study-1, yet they were not exactly the same; see **Figure 25.**

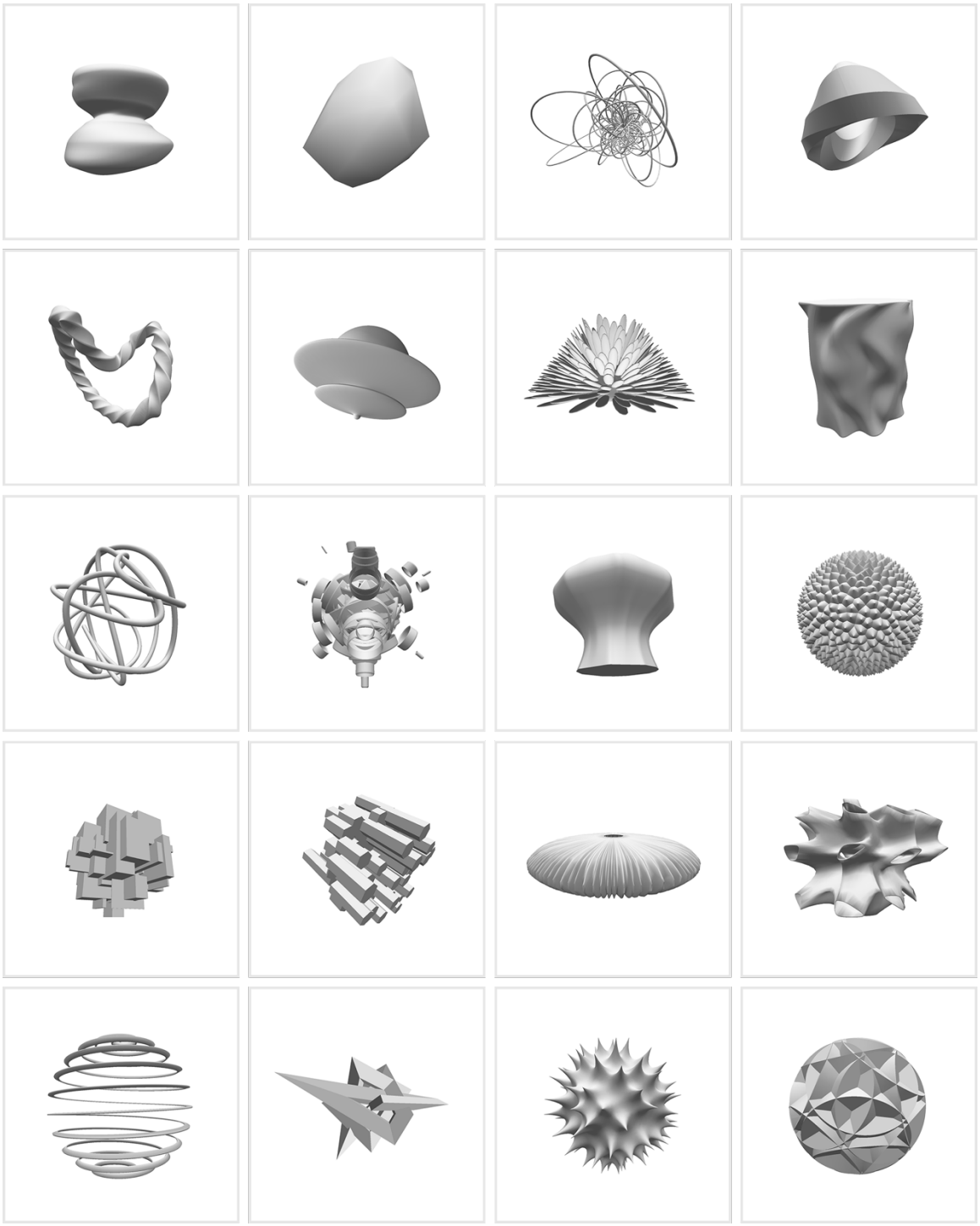


Figure 25. Screenshots of the stimuli used in study-2. These objects were digitally interactive; participants could move and rotate them. Compared to the prior study, these objects are similar but are not exactly the same, nor rendered exactly the same way.

The color of the objects was gray (specularMaterial = 70 brightness; normalMaterial = enabled; shininess = 10; colorMode = HSB). The background color was white (background = 100 brightness). The objects were illuminated using soft ambient light (ambientLight = 150 brightness) and three overhead lights (directly above, directionalLight = 0, 0, 40, 0, 100, 0; diagonally left, directionalLight = 0, 0, 40, -100, 100, -100; diagonally right, directionalLight = 0, 0, 40, 100, 100, -100).

4.3.2.3 Procedure

Each participant was randomly assigned one of the features to rate the novel objects. For example, participants assigned the moderate-extreme feature were asked "how moderate or extreme is this?" when they saw one of the stimuli, see **Figure 26**.

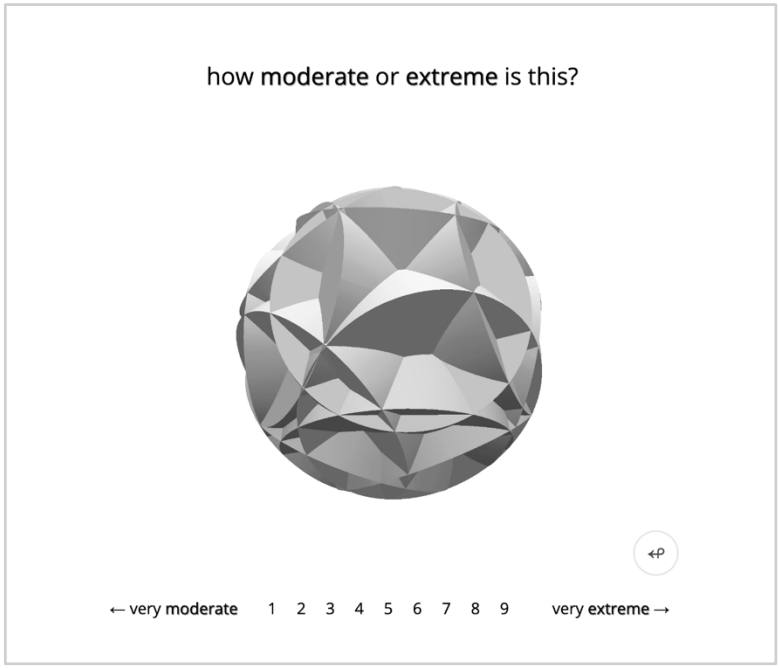


Figure 26. An example of a trial. Participants could rotate and zoom the object on their monitors.

A scale was shown after a five second delay after an object was presented. The delay was implemented to encourage participants to interact with the 3D objects. Once the scale was shown, participants were informed that they could submit their response. Participants responded by selecting a number on their keyboard from 1 through 9. The numbers represent a measure of meaning (Osgood et al., 1957), for example, 1 represents an impression that the object appears "moderate," 9 represents "extreme," and the numbers in-between represent an interpolation of meaning between "moderate" and "extreme". After submitting a response, the next object was presented. The order of the objects was randomized. After all the objects were evaluated, participants were asked to take a short break of any duration. Then they were asked to rate the objects once more.

Interacting with 3D objects within a web-browser at this time may not be a common experience. We therefore included a tutorial of how to interact with the objects along with the instructions for the task. Participants were shown how they could use their mouse to drag to rotate and scroll to zoom. On their screen, a prominent button, featuring a left arrow with a loop (↺), allowed participants to reset the objects to how they first appeared. After the objects were evaluated, demographics were collected along with any comments on the task. Most of the comments expressed enjoyment; none expressed any confusion as to how to manipulate the 3D objects.

4.3.2.4 Measures

4.3.2.4.1 Consistency

Consistency measures the reliability of an individual's impression each time they evaluate the same stimulus. Objects in this study were presented twice and consistency is

the Pearson's *r* correlation between each person's first and second set of impressions. To test the significance of each feature's consistency, Pearson's *r* was converted to Fisher's *z*. In this process, $r = 1$ becomes $z =$ infinity. Two participants had perfect consistency, $r = 1$, and therefore were held out of these analyses and later reincluded in the subsequent analyses. Using the *z* scores, we ran a series of two-tailed *t*-tests against zero to test for significance.

4.3.2.4.2 Consensus

Consensus measures how similar impressions are across people. We calculate this measure by using the average impression each person has of each object. Then in a leave-one-out process (e.g., Kurosu & Todorov, 2017), a participant's set of average impressions is correlated to the set of impressions averaged over everyone else. In other words, impression consensus represents how similar a person's impressions are to the norm. To ensure that the norm is informed by participants who are engaged in the task and are providing personally meaningful responses, we excluded participants with low-consistency ($r < .1$). The resulting analyses represents 1619 participants.

As with consistency, consensus uses Pearson's *r* correlation and so the coefficients range from -1 to 1. This process generates a measurement of consensus for every participant and represents the extent to which each participant's impressions were shared by others. If a participant's measured consensus is closer to -1, this suggests that their impressions are the opposite of the norm. A value closer to 0 would suggest the absence of a linear relationship. And a value closer to 1 suggests very similar impressions. The purpose of measuring consensus is to evaluate the reliability of each

feature in terms of how similarly it is evaluated across people. If a feature yields significantly low consensus (i.e., not significantly different from zero), that would suggest that its link to shapes is not reliably shared across people.

Looking ahead and examining the data, we tested its normality and the homogeneity of variance and both resulted in violations of assumptions for ANOVA. We therefore did not calculate ANOVA to compare the levels of consensus across features. We used a two-sided *t*-test against zero to determine if consensus for a feature was significant.

4.3.2.4.3 Consistency & Consensus

The relationship between the two measures of reliability (consistency, consensus) is analyzed using Pearson's *r* correlation using the data representing 1714 participants. This is the same dataset used to calculate the aggregate statistics of consistency; it excludes two participants who had perfectly consistent impressions. To calculate this correlation, we first convert consistency and consensus from *Pearson's r* to *Fisher's z* and then average their values across participants for each feature. This generates two scores for each feature: one representing consistency and the other representing consensus. We then correlate the two measures. Note that significance tests of the correlation between *Fisher z* scores have a degree of freedom equal to $n - 2$. Our hypothesis is that the two measures are significantly correlated and that features which are easier to engage in consistently by an individual, are also engaged in similarly across people.

4.3.2.4.4 Correlation between features

After examining the results of our data, our measures of reliability identified only one feature, unconstrained v. constrained, which did not elicit significant consensus. Because the overarching goal is to model shared object impressions, it was therefore crucial to exclude this feature. The resulting data represent 1485 participants and 56 features.

Before modeling the data, we describe the relationships between features using a correlational matrix. The matrix represents *Pearson's r* correlations as calculated by using the 'cor' function in R. The matrix is organized so that more highly correlated features are positioned more closely according to an underlying hierarchical cluster analysis of the pairwise correlation between features. The hierarchical cluster analysis was conducted using the same function in study-1. We report a plot visually conveying how similar features are engaged in while evaluating novel objects.

4.3.2.4.5 PCA

To model the data, we conducted an exploratory principal component analysis (PCA) using the 'principal' function in the 'psych' package (version 1.9) of R. This is a commonly accepted procedure for modeling impressions (e.g., Aaker, 1997; Oosterhof & Todorov, 2008). To briefly summarize this procedure, PCA projects all the pairwise relationships between features in a single multidimensional space. The number of dimensions (i.e., principal components) in the space is specified using Kaiser's rule, where in which dimensions with eigenvalues greater than or equal to 1 are kept (Kaiser, 1960). The data used to calculate PCA was the same dataset used to calculate the correlational matrix described in the section above; $n = 1485$.

While reporting the model's statistics (e.g., eigenvalues, proportion of variance explained), we focus on a subset of the 56 features. The subset was selected based on how the features most strongly correlated with a dimension (i.e., loading). When a dimension was strongly correlated with multiple features, we selected three features to add to the subset while also considering the frequency of their appearance in the free association data from study-1. We report the average absolute value of the loadings across the subset of features for each dimension, and also plot the relationship between the subset of features and each dimension.

To facilitate a discussion of the model's dimensions, each one was labeled using a noun. The intention for doing so was to fuel a discussion about how different impressions may relate to one another. To generate the labels, we considered how different features which load strongly on to the same dimension are relatable. The labels reflect broader jargon found in psychological research when multiple features load strongly onto a single dimension. When only one feature loaded strongly onto a dimension, the nominal version of one of the endpoints of that feature was used as the label.

4.3.2.4.4 Note about figures

All aggregate statistics on correlations were performed on *Fisher's z* transformations of *Pearson's r* —as already mentioned above. Just the same, figures plot *z* to *r* transformed values; error bars in the figures represent the standard error estimated using *z* scores and transformed to *r* values.

4.3.3 Results

4.3.3.1 Consistency

All 57 features were reliable in terms of consistency; see **Figure 27**. A two-tailed *t*-test against zero shows that impressions on average are significantly consistent when compared to chance, $t(56) = 39.54$, $p < .001$, $d = 5.24$, $r_M = .80$. Examining each feature independently, using a series of planned *t*-tests, further confirms that object impressions, regardless of the feature being examined, elicit significant consistency; see **Table 4**. The most consistent feature while evaluating novel objects was safe v. dangerous, $r_M = .85$. The least consistent feature was uncolored v. colored, $r_M = .46$.

Since the purpose of this paper is to model *shape*S*pace* using a data-driven approach, our tests of reliability (consistency, consensus) serve to not only describe data, but also filter out noise. All 57 features resulted in significant consistency on average, yet a few individuals were not consistent. We set a conservative threshold ($r < .1$) to remove outliers before proceeding to analyze consensus. Refer to the method section for more detail.

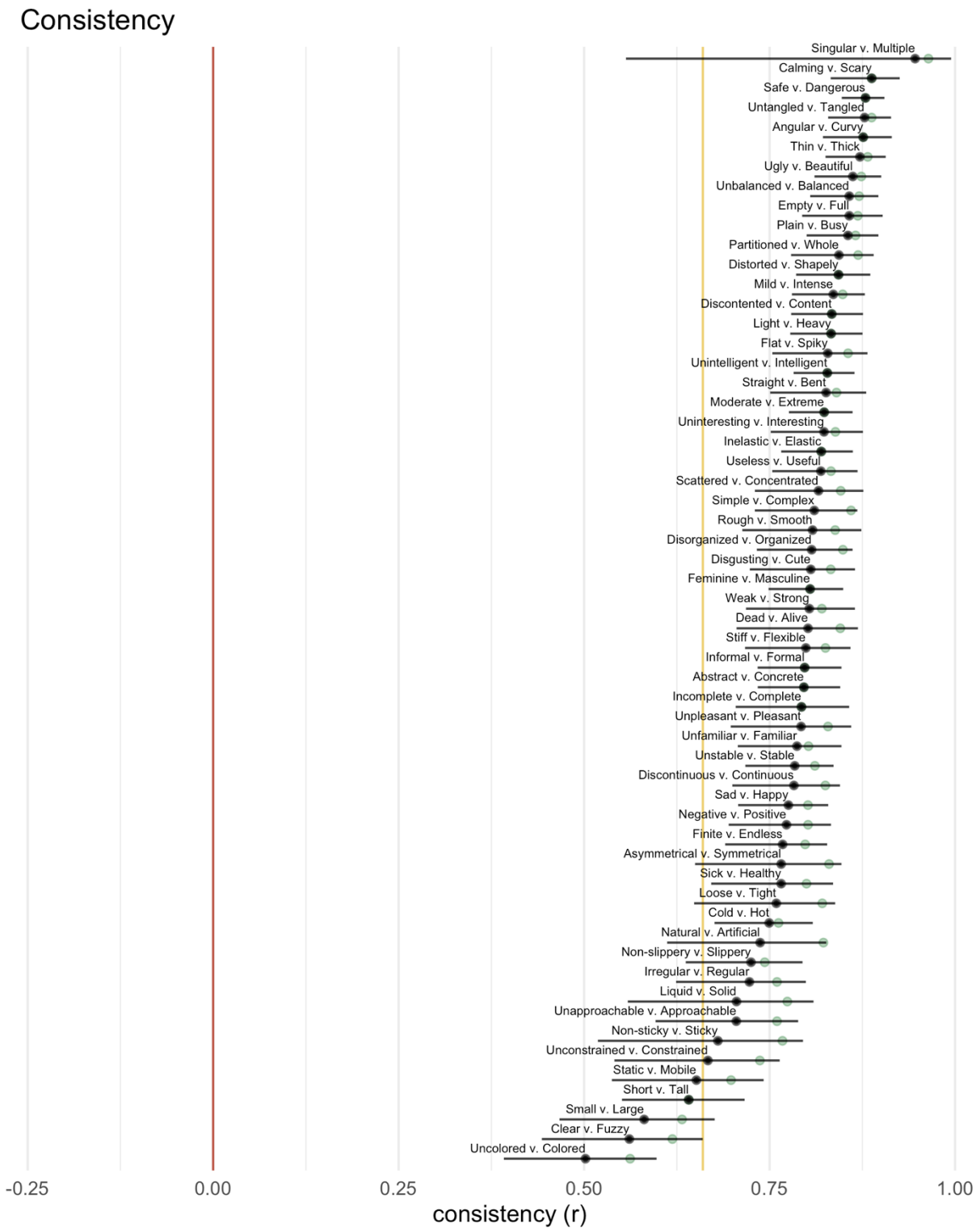


Figure 27. Consistency of each dimension. Gold line (r = .66), representing level of consistency found on average for novel object impressions (study-1, Kurosu & Todorov, 2017). Green points represent the average consistency of participants with good consistency (r > .1).

Table 4. Consistency for 57 features.

(Study-2) Consistency

Feature	*df*	*t*	*p*	*d*	r_M
Abstract v. Concrete	29	14.14	< .001	2.58	.80
Angular v. Curvy	30	13.55	< .001	2.43	.88
Asymmetrical v. Symmetrical	28	8.41	< .001	1.56	.77
Calming v. Scary	29	12.86	< .001	2.35	.89
Clear v. Fuzzy	28	7.85	< .001	1.46	.56
Cold v. Hot	30	12.67	< .001	2.28	.75
Dead v. Alive	29	9.59	< .001	1.75	.80
Discontented v. Content	29	14.94	< .001	2.73	.83
Discontinuous v. Continuous	30	11.11	< .001	2.00	.78
Disgusting v. Cute	29	10.95	< .001	2.00	.81
Disorganized v. Organized	28	11.98	< .001	2.23	.81
Distorted v. Shapely	29	14.12	< .001	2.58	.84
Empty v. Full	29	12.48	< .001	2.28	.86
Feminine v. Masculine	29	15.38	< .001	2.81	.80
Finite v. Endless	30	11.97	< .001	2.15	.77
Flat v. Spiky	30	11.50	< .001	2.07	.83
Incomplete v. Complete	28	10.39	< .001	1.93	.79
Inelastic v. Elastic	29	15.51	< .001	2.83	.82
Informal v. Formal	29	13.88	< .001	2.53	.80
Irregular v. Regular	28	9.84	< .001	1.83	.72
Light v. Heavy	30	14.89	< .001	2.67	.83
Liquid v. Solid	30	6.97	< .001	1.25	.71
Loose v. Tight	29	8.79	< .001	1.60	.76
Mild v. Intense	30	14.64	< .001	2.63	.84
Moderate v. Extreme	28	17.27	< .001	3.21	.82
Natural v. Artificial	28	7.96	< .001	1.48	.74
Negative v. Positive	29	11.85	< .001	2.16	.77
Non-slippery v. Slippery	29	10.90	< .001	1.99	.72
Non-sticky v. Sticky	29	6.37	< .001	1.16	.68
Partitioned v. Whole	29	12.70	< .001	2.32	.84
Plain v. Busy	31	14.03	< .001	2.48	.86
Rough v. Smooth	29	9.66	< .001	1.76	.81
Sad v. Happy	29	13.39	< .001	2.44	.78
Safe v. Dangerous	29	21.41	< .001	3.91	.88
Scattered v. Concentrated	29	10.41	< .001	1.90	.82
Short v. Tall	29	10.63	< .001	1.94	.64
Sick v. Healthy	29	10.06	< .001	1.84	.77
Simple v. Complex	31	11.13	< .001	1.97	.81
Singular v. Multiple	29	3.02	.005	.55	.95
Small v. Large	29	8.24	< .001	1.50	.58
Static v. Mobile	29	8.61	< .001	1.57	.65
Stiff v. Flexible	29	11.06	< .001	2.02	.80
Straight v. Bent	29	11.47	< .001	2.09	.83
Thin v. Thick	29	15.73	< .001	2.87	.87

Ugly v. Beautiful	29	14.73	< .001	2.69	.86
Unapproachable v. Approachable	30	9.05	< .001	1.63	.71
Unbalanced v. Balanced	28	14.68	< .001	2.73	.86
Uncolored v. Colored	29	7.85	< .001	1.43	.50
Unconstrained v. Constrained	28	7.90	< .001	1.47	.67
Unfamiliar v. Familiar	29	11.47	< .001	2.09	.79
Unintelligent v. Intelligent	29	17.87	< .001	3.26	.83
Uninteresting v. Interesting	29	12.01	< .001	2.19	.82
Unpleasant v. Pleasant	28	9.81	< .001	1.82	.79
Unstable v. Stable	29	13.50	< .001	2.46	.78
Untangled v. Tangled	30	14.68	< .001	2.64	.88
Useless v. Useful	29	13.07	< .001	2.39	.82
Weak v. Strong	29	10.62	< .001	1.94	.80

Notes. Planned two-tailed t-test against zero; d = Cohen's effect size; r_M = average Pearson's correlation. Arranged in alphabetical order of the left endpoint.

4.3.3.2 Consensus

People on average shared significantly similar impressions, $t(56) = 20.25$, $p < .001$, $d = 2.68$, $r_M = .79$. Closer examination of consensus for each feature using a series of *t*-tests reveals one feature, unconstrained v. constrained, did not elicit significant consensus $t(24) = 1.71$, $p = .10$, $d = .34$, $r_M = .08$. See **Table 5** for a list of consensus for all 57 features; **Figure 28** plots the values. The feature eliciting the highest consensus was safe v. dangerous, $r_M = .85$. Overall, the results of analyzing consensus shows that 56 features produce reliable impressions of novel objects.

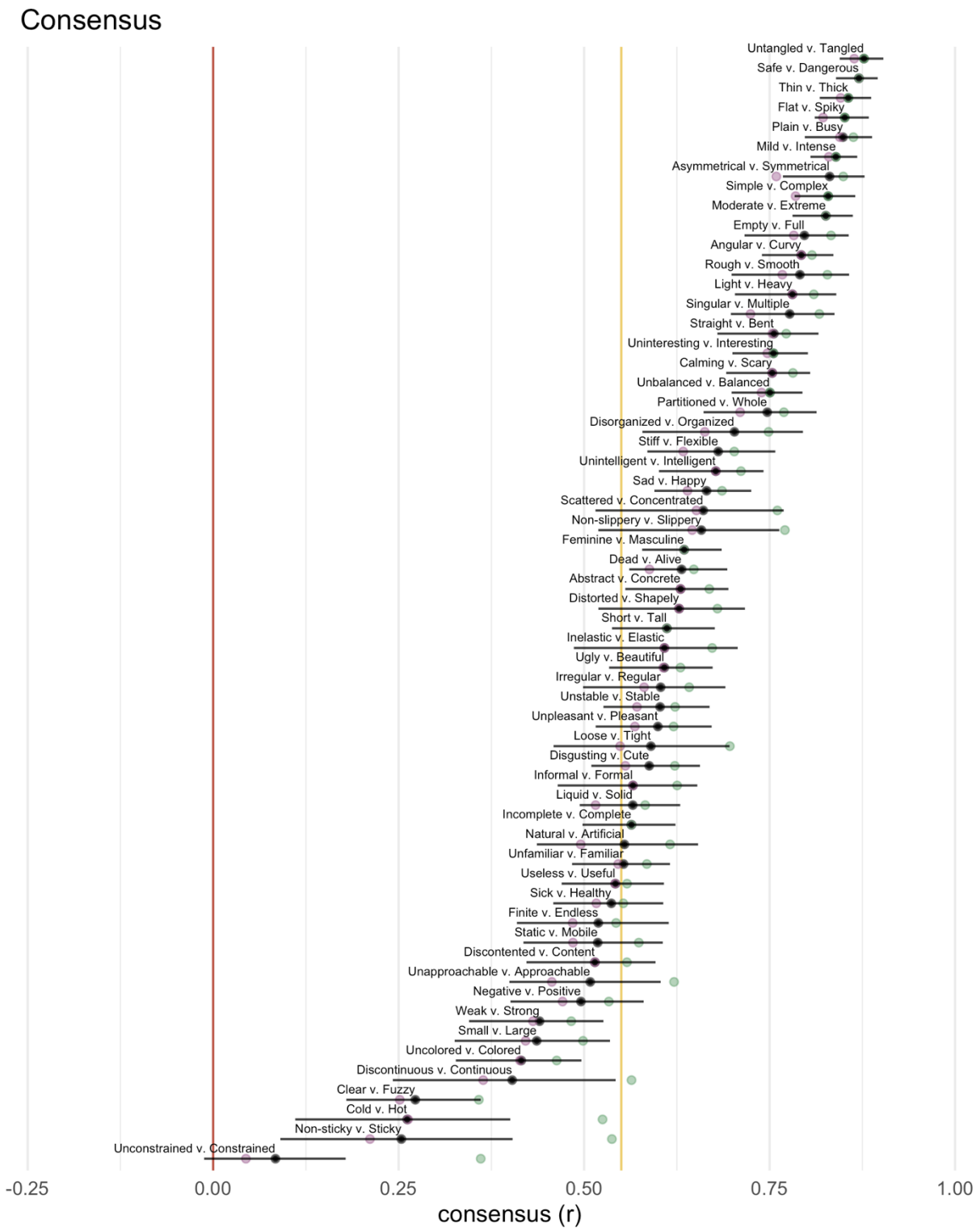


Figure 28. Consensus in order of magnitude. The red line represents zero consensus. The yellow line represents the average consensus previously seen with novel objects (r = .55, study-1, Kurosu & Todorov, 2017). Violet points represent mean consensus when participants with poor consistency (r < .1) are left

included in the analysis. Green points represent the recalculated mean consensus after excluding participants who shared low agreement ($r < .1$).

Table 5. Consensus for 57 features.

Feature	*df*	*t*	*p*	*d*	r_M
Abstract v. Concrete	29	12.62	< .001	2.30	.63
Angular v. Curvy	30	16.40	< .001	2.95	.79
Asymmetrical v. Symmetrical	24	13.28	< .001	2.66	.83
Calming v. Scary	29	14.75	< .001	2.69	.75
Clear v. Fuzzy	26	5.59	< .001	1.07	.27
Cold v. Hot	29	3.35	.002	.61	.26
Dead v. Alive	26	13.27	< .001	2.55	.63
Discontented v. Content	29	9.41	< .001	1.72	.51
Discontinuous v. Continuous	27	4.65	< .001	.88	.40
Disgusting v. Cute	27	11.81	< .001	2.23	.59
Disorganized v. Organized	25	8.06	< .001	1.58	.70
Distorted v. Shapely	29	8.88	< .001	1.62	.63
Empty v. Full	28	11.23	< .001	2.08	.80
Feminine v. Masculine	29	16.38	< .001	2.99	.63
Finite v. Endless	28	8.04	< .001	1.49	.52
Flat v. Spiky	28	18.80	< .001	3.49	.85
Incomplete v. Complete	29	13.62	< .001	2.49	.56
Inelastic v. Elastic	29	7.91	< .001	1.44	.61
Informal v. Formal	29	9.08	< .001	1.66	.57
Irregular v. Regular	26	9.08	< .001	1.75	.60
Light v. Heavy	30	11.82	< .001	2.12	.78
Liquid v. Solid	26	12.62	< .001	2.43	.57
Loose v. Tight	25	7.30	< .001	1.34	.59
Mild v. Intense	29	22.48	< .001	4.10	.84
Moderate v. Extreme	28	18.12	< .001	3.37	.83
Natural v. Artificial	23	7.80	< .001	1.59	.55
Negative v. Positive	27	8.95	< .001	1.69	.50
Non-slippery v. Slippery	28	7.22	< .001	1.34	.66
Non-sticky v. Sticky	24	3.02	.006	.60	.25
Partitioned v. Whole	27	11.06	< .001	2.09	.75
Plain v. Busy	30	15.27	< .001	2.74	.85
Rough v. Smooth	27	10.08	< .001	1.90	.79
Sad v. Happy	27	13.45	< .001	2.54	.67
Safe v. Dangerous	29	22.68	< .001	4.14	.87
Scattered v. Concentrated	27	6.95	< .001	1.31	.66
Short v. Tall	29	12.61	< .001	2.30	.61
Sick v. Healthy	27	11.28	< .001	2.13	.54
Simple v. Complex	27	17.86	< .001	3.38	.83
Singular v. Multiple	26	11.63	< .001	2.24	.78
Small v. Large	26	7.07	< .001	1.36	.47
Static v. Mobile	26	8.76	< .001	1.69	.52
Stiff v. Flexible	27	10.16	< .001	1.92	.68
Straight v. Bent	28	12.23	< .001	2.27	.76

Thin v. Thick	28	19.43	< .001	3.61	.86
Ugly v. Beautiful	28	12.50	< .001	2.32	.61
Unapproachable v. Approachable	27	7.98	< .001	1.51	.51
Unbalanced v. Balanced	27	17.53	< .001	3.31	.75
Uncolored v. Colored	26	8.46	< .001	1.63	.42
Unconstrained v. Constrained	24	1.71	.100	.34	.08
Unfamiliar v. Familiar	28	12.83	< .001	2.38	.55
Unintelligent v. Intelligent	29	12.43	< .001	2.27	.68
Uninteresting v. Interesting	28	16.35	< .001	3.04	.76
Unpleasant v. Pleasant	26	11.12	< .001	2.14	.60
Unstable v. Stable	27	12.17	< .001	2.30	.60
Untangled v. Tangled	29	21.11	< .001	3.85	.88
Useless v. Useful	28	12.18	< .001	2.26	.54
Weak v. Strong	28	8.23	< .001	1.53	.44

Notes. Planned two-tailed t-test against zero; d = Cohen's effect size; r_M = average Pearson's r correlation. Features arranged in alphabetical order by left endpoint.

4.3.3.3 Consistency & Consensus

The extent by which object impressions elicit consistency and consensus varies across different features; see **Figure 29**. However, the two measures of reliability are significantly correlated, $t(55) = 6.99$, $p < .001$, $r = .69$. As consistency increased for a feature, so too did consensus. The primary exception being 'unconstrained v. constrained,' which resulted in significant consistent, but not significant consensus. These results reinforce the importance of examining both measures of reliability separately; significant impression consistency does not guarantee significant impression consensus.

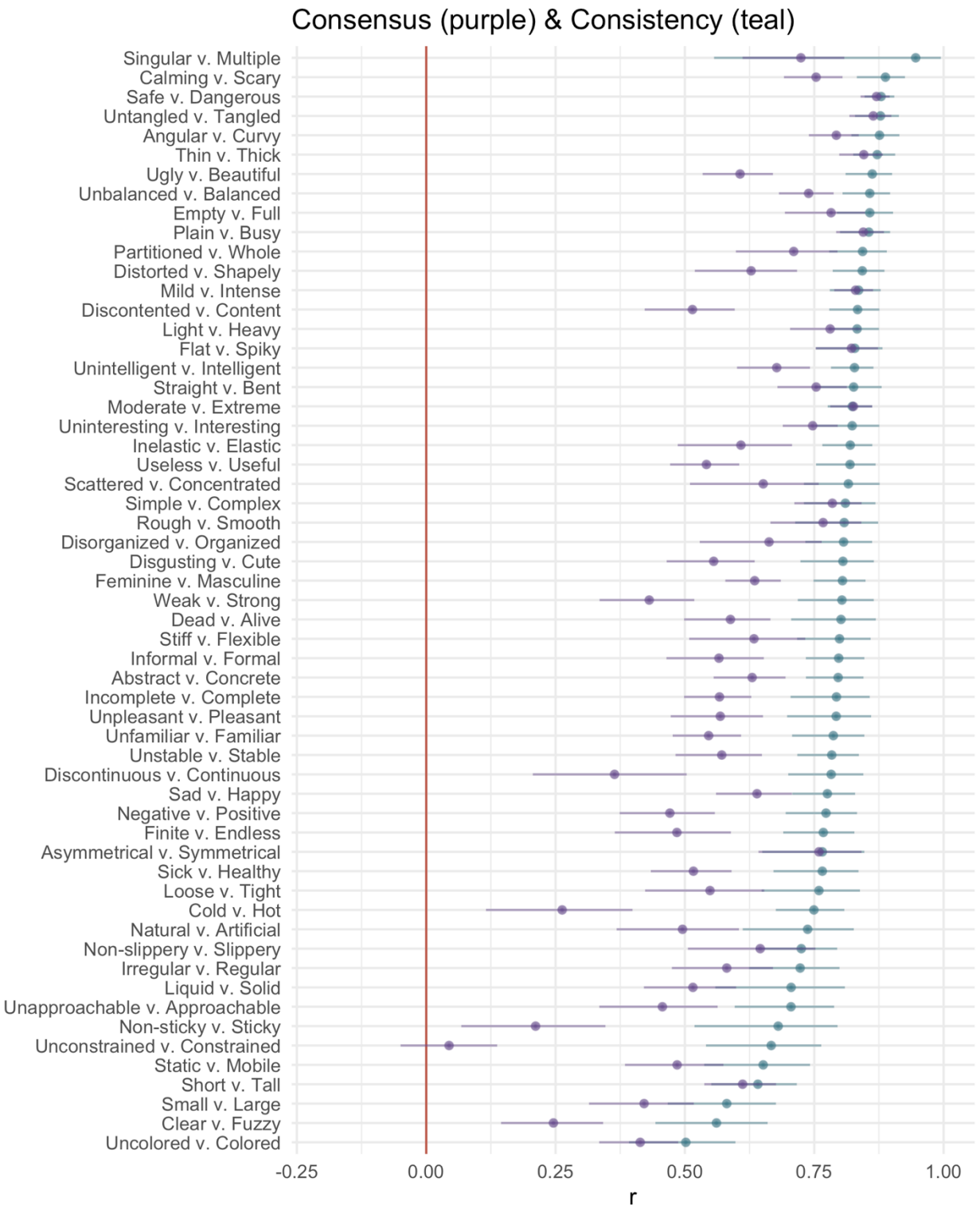


Figure 29. Consistency and consensus for study-2. In order of consistency magnitude. The red line represents zero consistency or consensus.

4.3.3.4 Correlation between features

To examine the relationship between features, we report every pairwise correlation in a matrix; see **Figure 30**. The result shows clusters of related variance. This suggests that shapes can systematically relate to impressions in more ways than capturable in one-dimensional space; exactly how many is the next question.

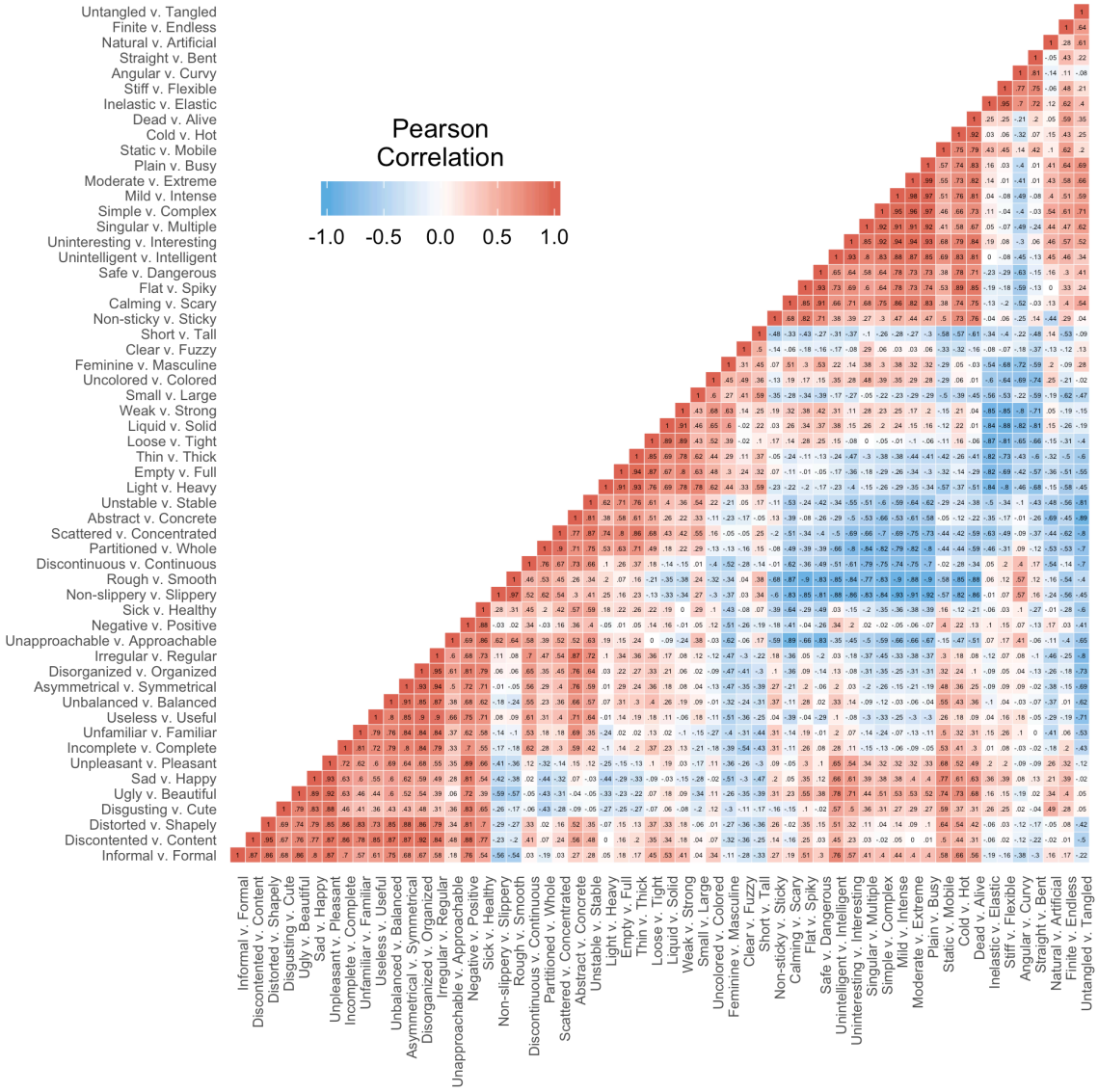


Figure 30. Correlation matrix arranged according to clusters. Red represents positive correlations. Blue represents negative correlations.

4.3.3.5 PCA

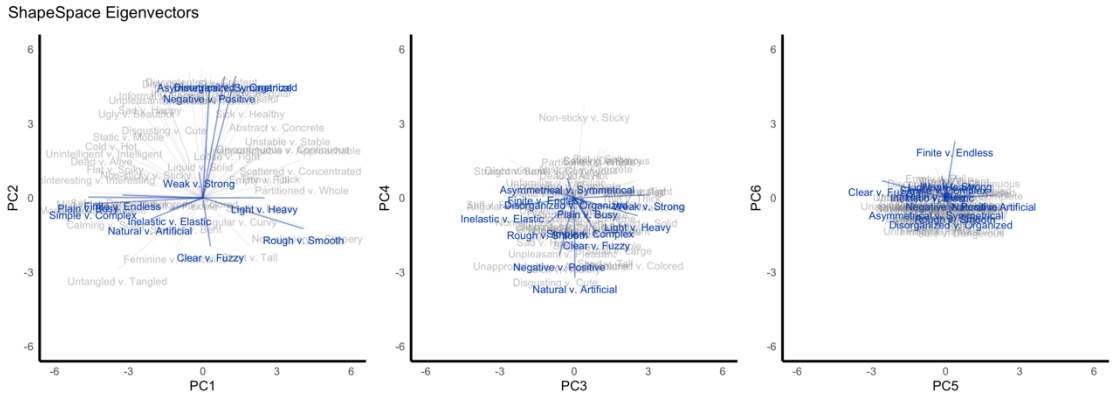


Figure 31. A projection of each of the 56 eigenvectors onto each of the six dimensions (PCs). A subset of 12 features are highlighted in blue. Longer and more parallel lines (parallel with a dimension) suggests a stronger relationship between a feature and a dimension. The vectors become progressively smaller from left to right as the eigenvalue decreases.

The results of the exploratory PCA shows that shared novel object impressions efficiently compress into a six-dimensional space; see **Figure 31** for a plot of the eigenvectors. Each of the six dimensions has been labeled as follow: *activity, fluency, intuitive physics, naturalness, clearness, endlessness*; see **Table 6**. The first three dimensions explains the majority (79%) of the variance and are discussed using a subset of three features each. The average loading based on the subset of features germane to each dimension was very high, *activity*, r = .92; *fluency*, r = .90; *intuitive physics*, r = .88. The last three dimensions explain a minority (14%) of the variance. With these dimensions, we focus on a single feature with a high loading each; *naturalness* = 'natural v. artificial', r = .66; clearness = 'clear v. fuzzy', r = .70; 'finite v. endless', r = .50. For a full list of the 56 features as they relate to each dimension see **Table 7.** All six dimensions together explain 93% of the variance among shared impressions of novel objects.

Table 6. The six-dimensional PCA performance in compressing information regarding 56 features.

(Study-2) PCA descriptive label and statistics

	PC1	PC2	PC3	PC4	PC5	PC6
	Activity	Fluency	Intuitive Physics	Naturalness	Clearness	Endlessness
loading	.92	.90	.88	.66	.70	.50
eigenvalue	18.49	15.14	10.88	4.64	1.77	1.13
variance explained	.33	.27 (.60)	.19 (.79)	.09 (.88)	.03 (.91)	.02 (.93)

Notes. Loading represents an average of the absolute value of three features' loading for the first three PCs. For the later three PCs, loading is the direct reported loading of a single feature. In the row reporting variance explained, the cumulative variance explained is reported in parentheses.

Table 7. Loadings (i.e., correlations) for all 56 features in the six-dimensional PCA. Highlighted are specific loadings for a focused subset of 12 features.

(Study-2) Correlations between dimensions and features

	PC1	PC2	PC3	PC4	PC5	PC6
	Activity	Fluency	Intuitive Physics	Naturalness	Clearness	Endlessness
Abstract v. Concrete	-.63	.63	.14	-.33	.15	-.10
Angular v. Curvy	-.30	-.10	-.86	-.02	.07	.19
Asymmetrical v. Symmetrical	-.18	.93	-.09	-.16	.10	-.05
Calming v. Scary	.82	-.12	.36	-.39	.01	-.07
Clear v. Fuzzy	-.06	-.37	.28	.30	.70	.15
Cold v. Hot	.77	.48	.10	-.28	-.03	-.04
Dead v. Alive	.86	.37	-.05	-.28	.19	.03
Discontented v. Content	.01	.97	.00	.14	-.05	-.08
Discontinuous v. Continuous	-.68	.46	-.29	-.32	-.09	.22
Disgusting v. Cute	.33	.60	-.24	.60	-.11	.01
Disorganized v. Organized	-.28	.93	-.10	-.03	-.06	-.14
Distorted v. Shapely	.12	.96	.00	.04	-.05	-.07
Empty v. Full	-.48	.23	.75	-.14	.11	.27
Feminine v. Masculine	.23	-.38	.75	-.02	-.07	-.19
Finite v. Endless	.68	.02	-.35	-.07	-.10	.50
Flat v. Spiky	.74	.30	.34	-.42	.02	-.15
Incomplete v. Complete	-.03	.89	-.10	-.19	-.24	.12
Inelastic v. Elastic	.28	-.09	-.90	.08	.15	.10
Informal v. Formal	.35	.87	.23	.19	-.04	.05
Irregular v. Regular	-.37	.89	-.07	-.17	.12	-.08
Light v. Heavy	-.52	.00	.80	.15	.05	.20
Liquid v. Solid	.02	.32	.89	.11	-.24	-.02
Loose v. Tight	-.20	.40	.82	-.15	-.18	.19
Mild v. Intense	.95	.04	.25	.04	.09	-.05
Moderate v. Extreme	.97	-.01	.16	.07	.08	.02
Natural v. Artificial	.45	-.16	.01	.66	-.53	.02
Negative v. Positive	-.05	.84	-.19	.47	-.06	.02
Non-slippery v. Slippery	-.88	-.23	-.29	.11	-.09	.00
Non-sticky v. Sticky	.48	.26	.13	-.75	.23	.01

Partitioned v. Whole	-.84	.14	.18	-.39	-.19	.13
Plain v. Busy	.97	.01	.17	.04	.10	.11
Rough v. Smooth	-.85	-.24	-.34	.21	-.10	-.09
Sad v. Happy	.45	.76	-.33	.27	.06	.02
Safe v. Dangerous	.75	.05	.41	-.40	-.17	-.21
Scattered v. Concentrated	-.83	.29	.39	-.15	-.10	.15
Short v. Tall	-.40	-.37	.39	.45	.12	-.17
Sick v. Healthy	-.40	.73	-.10	.49	.04	-.07
Simple v. Complex	.93	-.04	.20	.20	.02	.17
Singular v. Multiple	.85	.00	.29	.30	.24	.04
Small v. Large	-.48	.01	.56	.34	.40	.01
Static v. Mobile	.63	.55	-.34	-.13	.02	.18
Stiff v. Flexible	.16	.02	-.92	-.04	.25	.10
Straight v. Bent	.14	-.15	-.81	-.31	-.08	.25
Thin v. Thick	-.57	.23	.72	-.09	-.05	.24
Ugly v. Beautiful	.56	.73	-.10	.21	-.14	.01
Unapproachable v. Approachable	-.68	.45	-.29	.47	.08	.03
Unbalanced v. Balanced	-.07	.92	.02	-.16	.17	.11
Uncolored v. Colored	.12	.04	.76	.46	.22	.11
Unfamiliar v. Familiar	-.07	.83	-.24	-.22	.14	-.19
Unintelligent v. Intelligent	.83	.43	.22	.18	-.03	-.02
Uninteresting v. Interesting	.93	.22	.06	.16	.05	.02
Unpleasant v. Pleasant	.34	.83	-.15	.37	-.04	.04
Unstable v. Stable	-.72	.52	.31	-.04	.13	.11
Untangled v. Tangled	.72	-.55	-.12	.15	-.17	.17
Useless v. Useful	-.30	.84	-.18	.01	.11	-.13
Weak v. Strong	.03	.20	.94	-.02	-.12	.20

Note. Correlations (Pearson's r) are loading using the 'principal' function in the R 'psych' package (version 1.9.12).

4.3.4 Discussion

The results suggest a common language for describing shapes. The range of meaning likely shared is captured by a six-dimensional *shapeSpace* with the following dimensions: *activity, fluency, intuitive physics, naturalness, clearness, endlessness*.

The first dimension, *activity,* references a dimension of semantic space sharing the same name (Osgood et al., 1957). In both *shapeSpace* and semantic space, *activity* is correlated with features such as simple v. complex, uninteresting v. interesting, hot v. cold, and calming v. scary. Other strongly associated features, yet unique to *shapeSpace,* are dead v. alive, mild v. intense, moderate v. extreme, and plain v. busy.

The second dimension, *fluency*, references features associated with the fluency hypothesis of aesthetics (Reber et al., 2004). For example, asymmetrical v. symmetrical, disorganized v. organized, negative v. positive, ugly v. beautiful, unfamiliar v. familiar, and unpleasant v. pleasant are all strongly correlated with the *fluency* dimension within *shapeSpace,* and represent variables considered by the fluency hypothesis. The gist of the fluency hypothesis is that certain qualities, such as symmetry and familiarity, make it easier to process stimuli and thereby gives rise to positive feelings such as beauty (Reber et al., 2004). Other strongly associated features include discontented v. contented, distorted v. shapely, incomplete v. complete, informal v. formal, irregular v. regular, sad v. happy, sick v. healthy, unbalanced v. balanced, and useless v. useful. Future research inspired by the fluency hypothesis could examine the potential relationship aesthetics has with these other features.

The third dimension, *intuitive physics*, references a domain of research sharing the same name. Also known as 'naïve physics,' this research studies people's impressions of movement, interactions, and the physical attributes of objects (e.g., Kubricht et al., 2017). Although we did not show any objects interacting, physical attributes were nonetheless reliably present among the types of impressions. Angular v. curvy, inelastic v. elastic, light v. heavy, liquid v. solid, loose v. tight, stiff v. flexible, and weak v. strong were mapping to shapes similarly and strongly correlated with the third dimension of *shapeSpace*. Cold v. hot, short v. tall, and unstable v. stable, are arguably physical attributes too. However, these features did not correlate highly with the *intuitive physics* dimension. These results suggest that not all physical attributes are informed by shapes in

a similar way and that future research within the domain of intuitive physics could examine how the underlying mechanisms forming these impressions may be different.

The remaining three dimensions were labeled in reference to a single feature associated with that dimension. The fourth dimension, *naturalness*, references natural v. artificial. This dimension may relate to prior research on human-likeness and the uncanny valley (e.g., Wang et al., 2015), the perception of man-made objects (e.g., Fei-Fei et al., 2007), or the aesthetics of landscapes (Sevenant & Antrop, 2010). The fifth dimension, *clearness*, references clear v. fuzzy. This dimension may relate to prior research on consciousness and attention (e.g., Guilford & Hackman, 1936). Finally, the sixth dimension, *endlessness*, references finite v. endless. This last dimension may relate to prior research on perceptual unity (e.g., Johnson, 1997), or research on feature integration and the pop-out effect (e.g., Treisman, 1986).

Our model of *shapeSpace* indeed generates many interesting questions for further research. Yet the validity of *shapeSpace* is uncertain. Therefore, in the following study we examine how well the six-dimensional structure of *shapeSpace* replicates among a different sample of participants evaluating different novel objects.

4.4 Study-3

4.4.1 Introduction

Study-2 resulted in a model of *shapeSpace* with six dimensions— *activity, fluency, intuitive physics, naturalness, clearness, endlessness.* In the present study, we examine the validity of this model by conducting a conceptual replication. The general

procedure remained the same, however we asked the participants in this study to evaluate objects belonging to the same shape-family, rather than different shape-families. The features considered were a subset previously culled in study-2. Rather than having each participant consider just one of them, participants in study-3 had to consider all 12.

The departure in methodological details from study-2 affords the present study with the ability to address three uncertainties about *shapeSpace's* validity. Using different participants and different stimuli, we first modeled all the impressions using a confirmatory PCA. The purpose was to examine the external validity of *shapeSpace's* six dimensions in capturing variance among impressions of new shapes. Then, in our second examination of *shapeSpace's* validity, we modeled each shape-family separately to examine how well *shapeSpace* captures variance among impressions of more similar looking novel objects. Finally, we examined *shapeSpace's* validity in representing each individual's set of impressions. Together, these three levels of examination provide a robust audit of *shapeSpace* and the notion of a multidimensional latent space representing shared impressions of novel objects.

4.4.2 Method

4.4.2.1 Participants

Data from 107 participants are analyzed in study-3. Participation requirements included: adult age, normal vision, fluency in English, no prior exposure to our stimuli. The mean age was 19.48 (*SD* = 1.38). 74 were female (35 male, 1 non-binary). 48 were white (31 Asian, 10 Black, 3 Hispanic, 4 Other, 14 Multiracial).

Participants were recruited from Princeton University and the surrounding community. They either received credit for a class or financial compensation for their participation. Unlike study-1 and study-2, study-3 was conducted offline since its experimental design required longer participation. We recruited participants to come into the lab where they were provided a comfortable setting that was quiet and private. 123 participants were originally recruited. 13 participants were excluded because there was zero variance in their impressions during a block of the study. 3 participants were also excluded because their impressions overall and regardless of feature, had poor consistency ($r < .1$).

Participants viewed novel objects belonging to one of six shape-families: BB1, BB2, EG1, EG2, SA1, SA2. An *a priori* power analysis was conducted using G*Power3 (Faul et al., 2007) to determine the number of participants to recruit. We specified a two-tailed test of a correlation of .71, alpha of .05, and with a power of .80. The results of the power analysis suggested we needed at least 13 participants evaluating each shape-family. The final data analyses represent the following samples: BB1, $n = 19$; BB2, $n = 19$; EG1, $n = 18$; EG2, $n = 17$; SA1, $n = 18$; SA2, $n = 16$.

4.4.2.2 Stimuli

The stimuli consisted of 210 novel objects belonging to six shape-families, see **Figure 32**. Participants were shown only one shape-family (i.e., 35 objects). The shape-families are different from those used in study-1 and study-2. These shape-families were created by three practitioners of design, architecture, and art; Birk Binnard, Ethan Gross, and Seghier Abdelaziz. These practitioners responded to a request to donate shape-

families for research purposes on an online forum (https://discourse.mcneal.com). The purpose of the research was not disclosed to them.

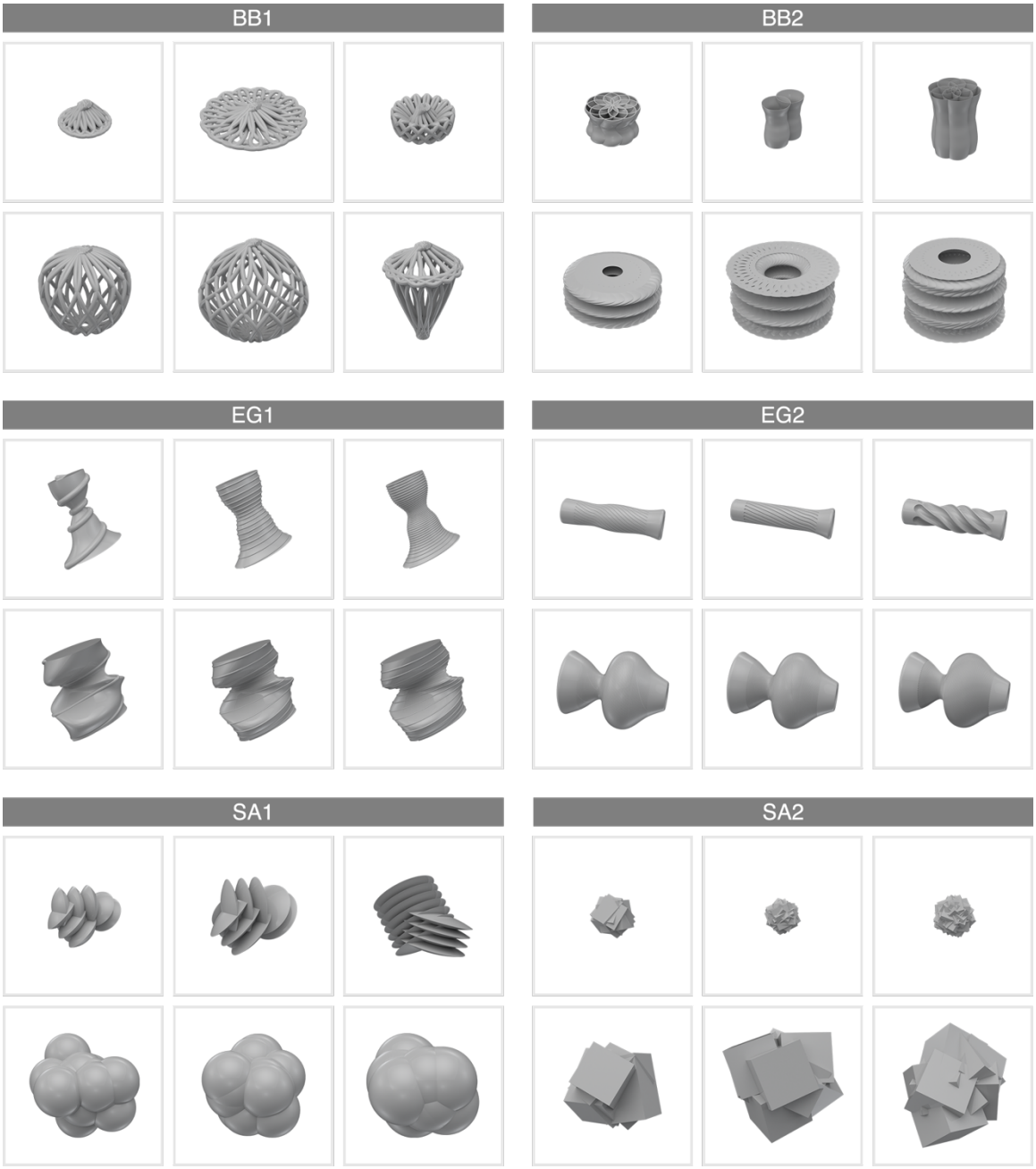


Figure 32. Example of study-3's stimuli. Six shape-families were represented across participants: BB1, BB2, EG1, EG2, SA1, SA2. Pictured are six objects sampled from each shape-family.

Using procedural generation, each object is defined by the values of shape-parameters established within each shape-family's algorithm. We manipulated three shape-parameters for each shape-family. Low-dispersed sampling was used to randomly generate values while also maximizing the diversity of shapes represented in each shape-family. This was accomplished by dividing each shape-parameter into five segments (initial, peninitial, middle, penultimate, ultimate), specifying certain combinations of those segments, and then randomly sampling values within each segment. There was a total of 35 combinations made. Of the total number of combinations, 27 were combinations of initial, middle, and last (ultimate) segments. The other 8 combinations were of peninitial and penultimate segments.

The objects were rendered as 2D images using V-Ray (version 3.60.03; www.chaosgroup.com). V-Ray is software which simulate physical illumination to create realistic photos of digital objects. The objects were rendered using a V-Ray material called "Plastic Simple Shiny Gray." The diffuse color of the shader was light gray (#BEBEBE) and roughness = 0. Other material settings were default values (reflection settings: color = #E9E9E9, glossiness = .98, Fresnel = on, IOR = 1.6, BRDF type = Blinn, max depth = 5, affect channels = color only; refraction settings: color = #000000, fog color = #FFFFFF, fog multiplier = 1, fog bias = 0, IOR = 1.6, glossiness = 1, affect shadows = on, refraction max depth = 5, affect channels = color only; opacity settings: opacity = 1, mode = stochastic). The illumination settings was set on 'high' (Raytrace noise limit = 0.01; antialiasing filter = 2, Lanczos; adaptive lights = 8; max trace depth = 5; max ray intensity = 20; primary rays = brute force; secondary rays = light cache; subdivs = 1500; sample size = 0.02; retrace = 4). Illumination was generated using the

program's default environment bitmap. The background of the objects was solid black (#000000). V-Ray camera took three 1000 x 1000 pixel images; camera settings: ISO = 800, F-number = 8, shutter speed = 100, vignetting = 0, vertical tilt = 0. The camera took a picture at three different positions (X, Y, Z): A = (55,55,55), B = (40, 66, 56), C = (25,72,57). At each position, the camera pointed towards the object located in the center of the space (0,0,0). Shape-families were scaled so that each constituent object would fit within the image boundaries. The three images provided participants with slightly different perspectives of each object. When viewed in temporal sequence, the images provided an illusion that the object was rotating (in depth).

4.4.2.3 Feature Selection

We tested 12 features culled in study-2; see **Table 19**. These features represented the six-dimensional structure of *shapeSpace.* When we projected the data from the 12 features into the 6-dimensional *shapeSpace*, 98% of the variance was explained; see the appendix for more detail.

Table 8. Replicating *shapeSpace* using a subset of 12 features.

Subset of 12 features representing 6 dimensions

Activity	Fluency	Intuitive Physics	Naturalness	Clearness	Endlessness
Plain v. Busy	Disorganized v. Organized	Weak v. Strong	Natural v. Artificial	Clear v. Fuzzy	Finite v. Endless
Rough v. Smooth*	Asymmetrical v. Symmetrical	Inelastic v. Elastic*			
Simple v. Complex	Negative v. Positive	Light v. Heavy			

*Notes. *feature was presented with flipped endpoints in study-3 relative to study-2.*

The six dimensions are: *activity, fluency, intuitive physics, naturalness, clearness, endlessness*. The 12 features are: asymmetrical v. symmetrical, clear v. fuzzy, disorganized v. organized, finite v. endless, inelastic v. elastic, light v. heavy, natural v. artificial, negative v. positive, plain v. busy, rough v. smooth, simple v. complex, weak v. strong. Each participant evaluated objects on the basis of all 12 features.

4.4.2.4 Procedure

The procedure of study-3 was similar to study-2 with a few differences. Participants were shown images of novel objects from one shape-family. There were 35 objects in a shape-family. Participants were tasked with rating each object, one at a time, according to the feature. For example, with regards to the feature "negative v. positive," participants were asked: "How negative or positive is this." After participants finished rating all the objects, they were offered a break before being asked to rate all the objects again on the basis of a different feature.

During each trial, participants rated the objects on a scale from 1 through 9. After the stimulus was presented, there was a 500ms delay before the scale appeared. The delay was designed to encourage meaningful engagement with the task. The objects were presented in a randomized order. The order in which the features appeared was randomized. After participants finished rating all the features and objects, they were asked to repeat the study. Because there were 35 objects, 12 features, and one repetition

of each trial, the total number of trials was 840. After participants finished these trials, demographic information was collected.



Figure 33. The presentation provided multiple rotations (in depth) of the objects.

Due to the length of this study, it was necessary to change the way the objects were presented. In study-2, we presented the objects as manipulatable 3D models. Here we presented the objects as three static images taken at slightly different rotations; see **Figure 33**. By clicking on buttons, participants could see the images from a different angle. If they clicked through the images in quick succession, the objects appeared to be animated and rotating in depth. Prior research suggests that rotated animations of objects provides greater clarity of an object's shape over just static images (Kourtzi & Shiffrar, 1999).

4.4.2.5 Measures

4.4.2.5.1 Across Shape-families (Consistency & Consensus)

First we attempted to replicate *shapeSpace* as discovered in study-2 by modeling all of study-3's data at once using a confirmatory six-dimensional PCA based on a correlation matrix. Just as how impressions in study-2 were made of objects belonging to

different shape-families, an analysis of all the data together represents impression of objects from different shape-families. We report the cumulative variance explained, factor congruence scores, and loadings. Factor congruence represents the similarity between the respective dimensions of the new confirmatory model and the original exploratory model of *shapeSpace*. Loadings again represent the correlation between features and dimensions. For the first three dimensions of *shapeSpace* loadings (*activity, fluency, intuitive physics*), since there are represented by three features each, we report their average of their absolute value.

The PCA loadings can be thought of as coordinates for orthogonal vectors in six-dimensional space. Doing so allows comparisons between models by comparing the similarity of these vectors in a measure called factor congruence (r_c). We compare the factor congruence between the new confirmatory model and the original model of *shapeSpace*. We use Procrustes transformation on the new confirmatory model's loadings before calculating factor congruence in order to test the best possible fit between models (e.g, McCrae et al., 1996). To do so, we use the 'Procrustes' function from the 'psych' package in R. To calculate factor congruence, we use the following equation:

$$r_c = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}$$

In this equation, x is associated with the new confirmatory model and y is associated with the original model of *shapeSpace*. Each variable represents the loadings of a single dimension. The original model of *shapeSpace* represents 56 features. However to calculate factor congruence, only the matching subset of 12 features was represented by y. The output ranges from -1 to 1; values closer to 1 represent a higher degree of

similarity between a particular dimension of both models. We report for each dimension: the factor congruence, loadings, eigenvalue, and the proportion of variance explained by each dimension.

Before modeling the data, we examined the reliability of each feature using measures of consistency and consensus. Greater detail regarding our calculation of consistency and consensus can be found study-2's method section.

4.4.2.5.2 Changes within Shape-family

The second level of examination divides the data by shape-family to examine impressions of objects that are more similar in shape. Each participant evaluated 12 features of one shape-family. Our measures are similar; however, the interpretation of the analyses is different because each model represents only one shape-family. Each model represents a test of how well the original *shapeSpace*'s structure generalizes to impressions of more similar shapes. If the factor congruence is high, the structure of *shapeSpace* arguably generalizes to any shared impression of a novel object. However, if the factor congruence scores are low or negative, the structure of a *shapeSpace* is arguable sensitive to the diversity of objects being evaluated together.

Because each shape-family is measured independently in this analysis, it was necessary to change the scope of our reliability measures as well. Consistency and consensus were analyzed separately for each feature and shape-family. At this level of analyses, not every feature of every shape-family resulted in significant consistency and consensus. Low levels ($r < .1$) were filtered out before calculating each PCA since the models represent averages. The measures of consistency and consensus based on

unfiltered data are reported in the results section. The potential implications of poor impression consistency or consensus at this level of analysis is explored in the discussion section. Additional statistics regarding the filtered data is reported in the appendix.

4.4.2.5.3 Individualized ShapeSpaces

The third level of examination involves modeling each participant's impressions separately using a confirmatory six-dimensional PCA. The purpose was to examine how well the original *shapeSpace* is represented by the impression of each participant. Our prior models involved averaging data across participants and an accompanying set of reliability measures (consistency, consensus). The combination of such reports implies that the models are representative of each participant. Due to the design of this study being that each participant evaluated all 12 features, we could model each participant's impressions individually to directly examine how well the structure of their impressions replicate the original model of *shapeSpace*.

At this level of analysis, additional measures of consistency and consensus were not necessary since the models do not represent averages across people. Likewise, no additional filtering was used; i.e., all 107 participants were modeled. After comparing these individual models to the original *shapeSpace*, we report the average statistics across the individualize models (e.g., loadings, eigenvalues, variance explained, factor congruence).

4.4.3 Results

4.4.3.1 Impressions Across Shape-families

The confirmatory six-dimensional model explains a similar amount of variance (92%) when compared to the prior study it attempts to replicate (study-2 = 93%), see **Table 9**. The first three dimensions captures the majority of the variance (75%), which is similar to what was observed in study-2 (79%). Thus, a six-dimensional space appears to reliably compress the variance among shared impressions of novel objects. However, the question remains whether or not the exact structure of the six dimensions is the same.

Table 9. Confirmatory model of ShapeSpace. The first three dimensions explains the most variance and replicates in meaning.

Study-3. Confirmatory PCA statistics

	PC1	PC2	PC3	PC4	PC5	PC6
	Activity	Fluency	Intuitive Physics	Naturalness	Clearness	Endlessness
loading	.79	.83	.78	.32	.66	.62
eigenvalue	4.96	2.45	1.59	.94	.63	.50
variance explained	.41	.20 (.62)	.13 (.75)	.08 (.83)	.05 (.88)	.04 (.92)
factor congruence	.86	.92	.90	.18	.68	.90

Notes. Loading is an average of the absolute value of a priori features for the first three dimensions, and a direct copy of a single feature for the later three dimensions. Statistics regarding the first three dimensions are bolded because they have an eigenvalue greater than one. In the row with variance explained, the cumulative value is showed in parentheses. Factor congruence measures the similarity between the exploratory model of study-2 and the present confirmatory model.

An examination of the loadings (r) and factor congruence (r_c) between the new confirmatory model and original model of *shapeSpace* shows that some of the dimensions, but not all, reliably represents the same meaning. The first three dimensions maintain a similarly strong relationship to their constituent features; *activity*, $r = .79$, , $r_c = .86$; *fluency*, $r = .83$, $r_c = .92$; *intuitive physics*, $r = .78$, $r_c = .90$. The later three dimensions, which explain a minority of the variance (15%), had mixed results. *Naturalness* had the lowest loading, $r = .32$, and factor congruence, $r_c = .18$; followed by

clearness, $r = .66$, $r_c = .68$; and *endlessness*, $r = .62$, $r_c = .90$. Thus, it appears that the first three dimensions may be more reliable across different people and shapes.

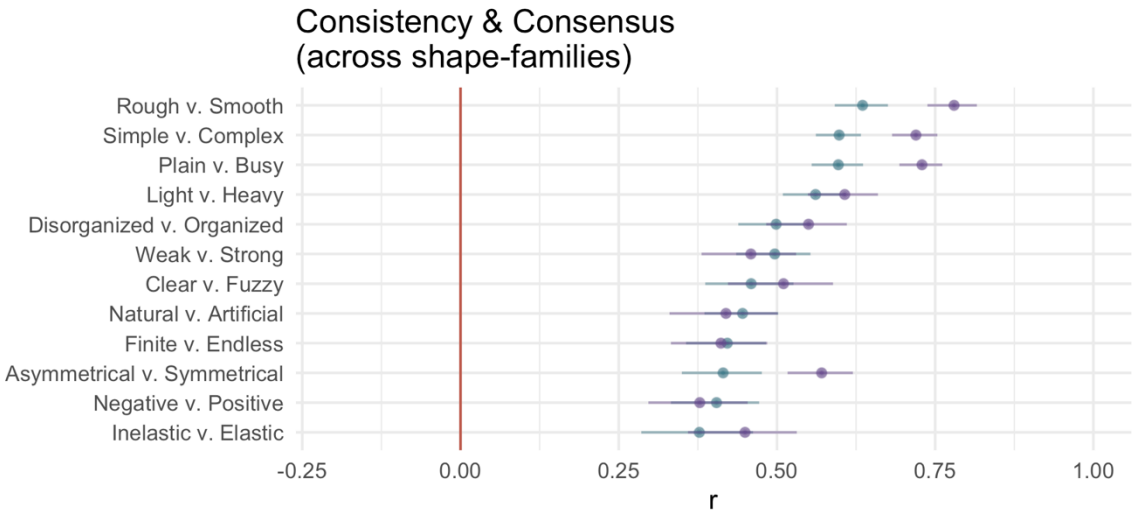


Figure 34. Reliability of impressions for each feature (across shape-families). Teal represents consistency, violet represents consensus. The whiskers represent a unit of standard error above and below. The red vertical line represents the baseline of zero. The values represent Pearson's *r* correlation after transforming back from Fisher's *z* scores.

Checking the reliability of each feature, there was significant impression consistency, $t(11) = 16.00$, $p < .001$, $d = 4.62$, $r_M = .50$, $r_{SD} = .12$; and consensus, $t(11) = 10.29$, $p < .001$, $d = 2.97$, $r_M = .56$, $r_{SD} = .21$. Examining the features separately in a series of planned *t*-tests against zero, consistency and consensus for each feature is indeed significant; for consistency, see

Table 10; for consensus, see **Table 11**; for a plot of both, see **Figure 34.** These results show that the features identified in study-2 can generalize to impressions of a different set of novel objects regardless of whether or not the exact structure of the original *shapeSpace* replicates perfectly.

Table 10. Consistency

(Study-3) Consistency across shape-families

Feature	*df*	*t*	*p*	*d*	r_M
Asymmetrical v. Symmetrical	106	11.33	< .001	1.09	.41
Clear v. Fuzzy	106	11.00	< .001	1.06	.46
Disorganized v. Organized	106	14.00	< .001	1.35	.50
Finite v. Endless	106	11.46	< .001	1.11	.42
Inelastic v. Elastic*	106	7.55	< .001	.73	.38
Light v. Heavy	106	17.15	< .001	1.66	.56
Natural v. Artificial	106	12.90	< .001	1.25	.45
Negative v. Positive	106	10.08	< .001	.97	.40
Plain v. Busy	106	21.38	< .001	2.07	.60
Rough v. Smooth*	106	20.86	< .001	2.02	.64
Simple v. Complex	106	24.38	< .001	2.36	.60
Weak v. Strong	106	13.69	< .001	1.32	.50

*Notes. Planned two-tailed t-tests against zero; d = Cohen's effect size; r_M = average Pearson's r correlation. *feature is flipped relative to study-2 when tested in study-3.*

Table 11 Consensus

(Study-3) Consensus across shape-families

Feature	*df*	*t*	*p*	*d*	r_M
Asymmetrical v. Symmetrical	106	16.62	< .001	1.61	.57
Clear v. Fuzzy	106	9.83	< .001	.95	.51
Disorganized v. Organized	106	13.27	< .001	1.28	.55
Finite v. Endless	106	9.43	< .001	.90	.41
Inelastic v. Elastic*	106	8.80	< .001	.85	.45
Light v. Heavy	106	15.75	< .001	1.52	.61
Natural v. Artificial	106	8.43	< .001	.82	.42
Negative v. Positive	106	8.51	< .001	.82	.38
Plain v. Busy	106	25.14	< .001	2.43	.73
Rough v. Smooth*	106	20.61	< .001	1.99	.78
Simple v. Complex	106	23.97	< .001	2.32	.72
Weak v. Strong	106	10.26	< .001	.99	.46

*Notes. Planned two-tailed t-tests against zero; d = Cohen's effect size; r_M = average Pearson's correlation. *feature is flipped relative to study-2 when tested in study-3.*

4.4.3.2 Impressions Within Shape-families

How well does a six-dimensional *shapeSpace* (as discovered in study-2) account for impressions of objects belonging to the same shape-family? Because objects sampled from the same shape-family, rather than different shape-families, results in smaller changes, the question here tests *shapeSpace* appropriateness for codifying impressions of subtle differences in shapes. We modeled impressions for each of the six shape-family separately, see **Table 12**.

Table 12. *ShapeSpace* performance in codifying impressions of objects belonging to the same shape-family.

(Study-3) Model of object impressions for each shape-family

		PC1	PC2	PC3	PC4	PC5	PC6
Shape-family		Activity	Fluency	Intuitive Physics	Naturalness	Clearness	Endlessness
BB1	loading	.76	.87	.80	.43	.27	.45
	eigenvalue	6.42	3.07	1.52	.26	.23	.15
	variance explained	.53	.26 (.79)	.13 (.92)	.02 (.94)	.02 (.96)	.01 (.97)
	factor congruence	.72	.85	.90	.50	.35	.77
BB2	loading	.73	.88	.85	.25	.25	.45
	eigenvalue	7.06	3.85	.42	.19	.15	.12
	variance explained	.59	.32 (.91)	.04 (.94)	.02 (.96)	.01 (.97)	.01 (.98)
	factor congruence	.73	.87	.81	.76	.67	.85
EG1	loading	.73	.86	.50	.57	.31	.59
	eigenvalue	7.50	2.16	1.30	.50	.18	.12
	variance explained	.63	.18 (.81)	.11 (.91)	.04 (.96)	.02 (.97)	.01 (.98)
	factor congruence	.75	.81	.75	.70	.68	.85
EG2	loading	.60	.57	.85	.04	.51	.42
	eigenvalue	8.96	2.21	.39	.10	.09	.06
	variance explained	.75	.18 (.93)	.03 (.96)	.01 (.97)	.01 (.98)	.01 (.98)
	factor congruence	.65	.44	.80	.50	.48	.71
SA1	loading	.86	.70	.36	.18	.50	.39
	eigenvalue	8.01	2.40	.81	.32	.17	.10
	variance explained	.67	.20 (.87)	.07 (.94)	.03 (.96)	.01 (.98)	.01 (.98)
	factor congruence	.56	.79	.85	.34	.57	.56
SA2	loading	.72	.67	.72	.27	.08	.34
	eigenvalue	9.54	1.31	.62	.17	.11	.07
	variance explained	.80	.11 (.90)	.05 (.96)	.01 (.97)	.01 (.98)	.01 (.99)

factor congruence	.66	.83	.74	.68	.23	.77
average factor congruency across shape-families	.68	.77	.81	.58	.50	.75

Notes. Loading is an average of the absolute value of a priori features for the first three dimensions, and a direct copy of a single feature for the later three dimensions. Statistics with an eigenvalue above one are emboldened. In the row with variance explained, the cumulative value is showed in parentheses. Factor congruence measures the similarity between the exploratory model of study-2 and each shape-family's model.

Six dimensions explain much of the variance among impressions of objects from the same shape-family; BB1 = 97%, BB2 = 98%, EG1 = 98%, EG2 = 98%, SA1 = 98%, SA2 = 99%. However, two dimensions explains much of the variance within four shape-families; BB2 = 91%, EG2 = 93%, SA1 = 87%, SA2 = 90%. For the other two shape-families, three dimensions were sufficient; BB1 = 92%, EG1 = 91%. This suggests that the variance in impressions found within each shape-family was more compressible than variance from across shape-families.

Factor congruence and loadings varied for each shape-family. Sometimes a dimension of *shapeSpace* replicated well with impressions for a specific shape-family. For example, the *intuitive physics* dimension of BB1 replicated study-2's *shapeSpace* well, $r = .80$, $r_c = .90$. However, the *clearness* dimension of BB1 did not replicate as well, $r = .27$, $r_c = .35$. Thus, six dimensions explain a lot of variance among impressions of objects that are similar in shape, however the structure of the six-dimensional model shifts somewhat in meaning.

Table 13. Consensus for each feature of each shape-family. Consensus varied in significantly depending on the feature and shape-family. While most impressions were consensual, a few (highlighted) did not result in significant levels of consensus.

(Study-3) Consensus for impressions within each shape-family

Dimension	Feature	BB1	BB2	EG1	EG2	SA1	SA2
Activity	Plain v. Busy	.72***	.72***	.71***	.60***	.74***	.87***

	Rough v. Smooth	.76***	.60***	.65***	.88***	.88***	.84***
	Simple v. Complex	.69***	.73***	.74***	.52***	.74***	.85***
Fluency	Disorganized v. Organized	.32***	.34**	.78***	.64***	.72***	.62***
	Asymmetrical v. Symmetrical	.59***	.55***	.82***	.43***	.58***	.55***
	Negative v. Positive	.53***	.18	.53***	.51***	.55**	.52**
Intuitive Physics	Weak v. Strong	.69***	.70***	.39***	.30	.26*	.46**
	Inelastic v. Elastic	.59***	.53***	.35*	.18	.84***	.46
	Light v. Heavy	.55***	.75***	.41***	.81***	.57***	.56***
Naturalness	Natural v. Artificial	.49**	.53***	-.09	.60**	.74***	.15
Clearness	Clear v. Fuzzy	.43**	.46**	.56***	.80***	.40**	.71**
Endlessness	Finite v. Endless	.33***	.56***	.48***	.44**	.14	.58*

*Notes. Level of consensus after filtering out participant data that with low consistency (r < .1) for a specific feature. Significance was tested on data before low levels of consensus (r < .1) was filtered out for the PCA. *p < 0.05, **p < .01, ***p < .001. The significant tests were planned two-tailed t-tests against zero using Fisher's z-scores.*

Examining the reliability of the data at the level of shape-family, not all features resulted in significant consensus; see **Table 13.** Here we report the levels of consistency and consensus before blocks of impressions with poor consistency and consensus were filtered, see Error! Reference source not found.**.** For example, a block of impressions representing the 'negative v. positive' feature and the BB2 shape-family did not have significant consensus, $t(16) = 1.44$, $p = .165$, $d = .35$, $r_M = .18$. Filtering out impressions with poor consistency and consensus avoids the influences of outliers in our models and ensures that the models represent only shared impressions.

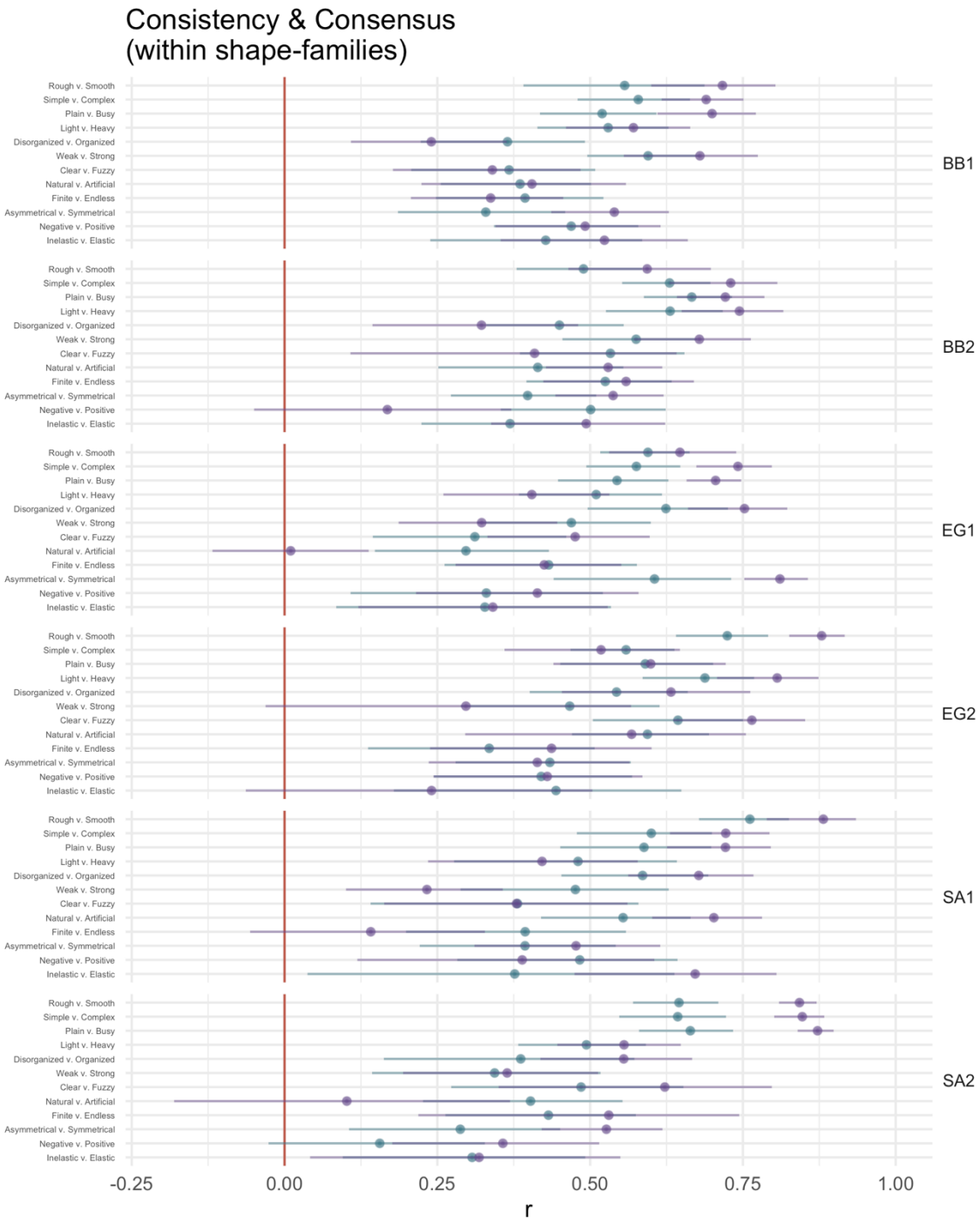


Figure 35. Consistency (teal) and consensus (purple) within each shape-family. (study-3) No data filtering was used in these calculations.

4.4.3.3 Individualized ShapeSpaces

We modeled each participant's impressions using a six-dimensional confirmatory PCA. The results provide further evidence that a six-dimensional structure explains the majority of variance reliably in impressions. In relation to the original *shapeSpace* from study-2, we discover evidence for a six-dimensional structure; see **Table 14**. The average cumulative variance explained by the individualized models was 90%. The first three dimensions explain the majority of the variance (75%).

Table 14. Summary statistics of individualized models of each participant's impressions. The values represent the average across all the individualized models.

(Study-3) Individualized six-dimensional confirmatory models

	PC1	PC2	PC3	PC4	PC5	PC6
	Activity	Fluency	Intuitive Physics	Naturalness	Clearness	Endlessness
loading	.64	.62	.60	.41	.53	.52
eigenvalue	5.71	2.13	1.11	.78	.59	.46
variance explained	.48	.18 (.66)	.09 (.75)	.07 (.81)	.05 (.86)	.04 (.90)
factor congruence	.70	.74	.75	.62	.62	.64

Notes. All of these values represent averages across individualized models. Correlation based PCA computed using the 'principal' function. Loading is an average of the absolute value of the first three PCs and a direct copy for the last three; it represents the correlation between the scores before and after PCA. Eigenvalues = "SS Loadings". Variance explained is the proportion of variance in the data captured by each dimension; the cumulative amount is in parentheses. Factor congruency scores are after Procrustes transformation; they measure the congruency of the same PC between study-2 and study-3.

Examining the structure of the individualized models while comparing them to the original *shapeSpace*, the results show that the structure's meaning shifts between participants, see **Figure 36**. The first three dimensions still tend to be less mutable, however the difference is attenuated. For example, the difference in loadings between the first three and last three dimensions was .13; in the original *shapeSpace* model the difference was .28.

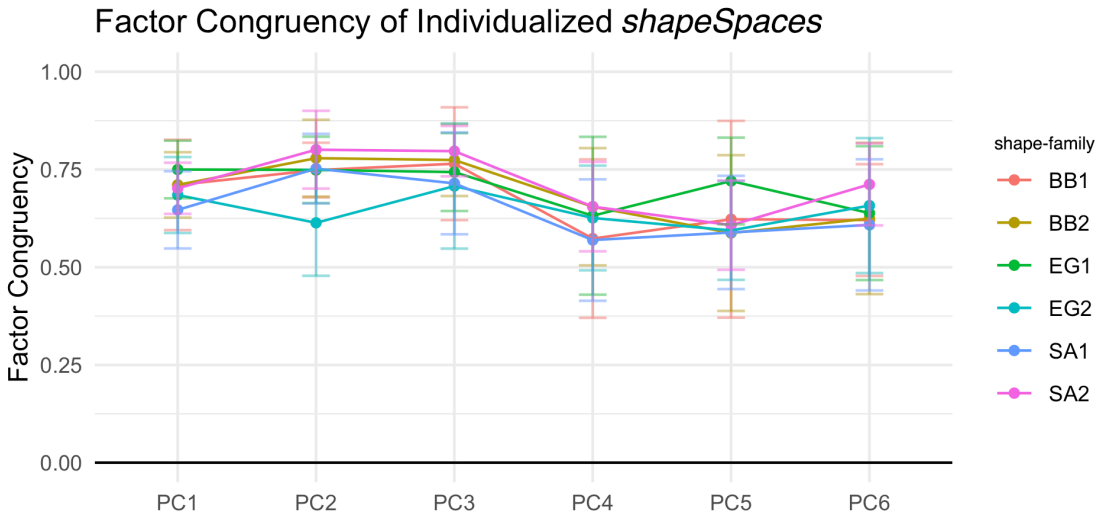


Figure 36. Factor congruency comparing individualized models to the original *shapeSpace* model are organized according to shape-family. Each participant saw objects belonging to one shape-family. The whiskers around the mean values represent one standard deviation. The six-dimensional structure shifts somewhat in meaning across participants.

4.4.4 Discussion

The results show that a confirmatory six-dimensional PCA captures much of the variance among shared impressions of novel objects. While comparing the structures between the confirmatory model and study-2's exploratory model, the meaning appears to shift somewhat; the last three dimensions shifting more so.

Future research could seek to more appropriately label the last three dimensions given their mutability. Our current research tested their validity using a limited number of features and it may be worth returning to the free association data to extract additional semantically related features. For example, adjectives related to 'clear v. fuzzy,' include fluffy and pokey. Adjectives related to 'natural v. artificial,' include organic and digital.

Adjectives related to 'finite v. endless,' include neverending and infinite. Alternatively, we previously speculated in the discussion of study-2 that these dimensions may be related to prior research on attention (e.g., Guilford & Hackman, 1936), unity (e.g., Johnson, 1997), and feature integration (e.g., Treisman, 1986). Therefore, prior research represents an alternative strategy for better understanding the later three dimensions.

Between individuals, it is possible that the last three dimensions are what primarily accounts for idiosyncratic differences across people's impressions, while the first three dimensions may help people to share them. Together, the congruent and incongruent dimensions of a person's own six-dimensional *shapeSpace* could explain how people are able to communicate descriptions of novel objects, and yet can sometimes find disagreement when talking to others.

In a famous experiment by Asch (1956), participants evaluated the lengths of lines after confederates expressed a false evaluation. Asch's experiments demonstrated the influence a majority has on an individual to conform to a group. If *shapeSpace* exists to facilitate socialization by way of communication, the mechanisms supporting its existence may also explain why Asch's participants 'felt' a need to conform. The same mechanisms may also explain why people are likely to expect their impressions to be shared by others (A. Kurosu & Todorov, 2017); i.e., false-consensus effect (Marks & Miller, 1987). The confederates in Asch's experiments may have been persuading a participant to shift their *shapeSpace* to accommodate a new mapping between stimuli and impressions. If this reframing of the situation is true, future research could reveal prerequisites and consequences of pressuring a person to shift their own *shapeSpace* to match the impressions of a group.

Over time, repeated exposures in the absence of contradictory feedback enhances impressions (e.g., Zajonc, 1968). We previously saw this effect when comparing impressions of novel objects between two exposures and ten; both consistency and consensus increased with more exposures (A. Kurosu & Todorov, 2017). Because participants are not sharing information or conspiring to agree, the rise in consensus may be attributable to people becoming more apt with navigating their own *shapeSpace* in order to map visual information to impressions. This again assumes that *shapeSpace* exists to facilitate socialization. But then why does the level of consensus significantly vary on the basis of an interaction between a feature and shape-family?

After dividing the data to model each shape-family separately we found varying levels of consistency and consensus. Poor consistency is likely attributable to testing fatigue. Participants may have been poorly engaged in the task because of the many trials and the subtle differences in stimuli. Respondent fatigue may also explain why several participants had to be excluded from this study's analyses due to a lack of variance across a portion of their responses. Longer tasks are more susceptible to such "straight-line" responses that diminish reliability (Herzog & Bachman, 1981). For future research, we recommend reducing the number of trials, or dispersing portions of the task across multiple sessions (e.g., Peytchev & Peytcheva, 2017).

With regard to poor consensus, our results replicate a prior finding that an interaction between feature and shape-family dictates the level of impression consensus (A. Kurosu & Todorov, 2017). The results showed that the variety of shapes represented by the stimuli affects consensus. While examining impressions across, and then within a shape-family, consensus was found and then lost among some features. This suggests that

the original *shapeSpace* may represent a general shape space that is engaged while looking across objects with diverse shapes. However, when looking across objects with similar shapes, people may engage a more specialized *shapeSpace*. This reasoning could explain why previously published multidimensional latent spaces of specific categories have different dimensional structures. For example, impressions of faces vary reliably by two dimensions: trustworthiness and dominance (Oosterhof & Todorov, 2008). Future studies examining how the mind shifts between representational spaces is warranted.

4.5 Chapter Discussion

The primary purpose of this chapter was to model shared impressions of novel objects in order to understand how they may form. In study-1 we collected free-association data to inform a set of potentially shared features. In study-2 we test the reliability of these features and model those which result in shared impressions. We create our first model of *shapeSpace* using an exploratory PCA. In study-3 we collect data from different participants responding to different shape-families. We then compute a series of confirmatory PCA modeling the data at different levels. The first level examined all the impressions across shape-families together in a close replication of study-2. The second level examined impressions of one shape-family at a time to probe how well *shapeSpace* explains variance among impressions of more similar looking shapes. The third level examines individuals one at a time to probe how well *shapeSpace* explains variance across a person's set of impressions.

Taken together, our results suggest that people develop a six-dimensional *shapeSpace* throughout their lives. We theorize that *shapeSpace* provides a basis for sharing experiences. Through conversations people are often trying to find common ground (Brennan & Clark, 1996). Through this process, an individual's *shapeSpace* may become entrained. This could explain why participants may form shared impressions of novel objects. If this is true, then *shapeSpace* is fundamental not only for communication, but also for 'shared reality' (for a review, see Echterhoff et al., 2009).

Examining the common structure of *shapeSpace* has implication for bridging research on different types of features. For example, we found that 'ugly v. beauty' and 'useless v. useful' both loaded onto the *fluency* dimension. This finding has not been recorded in the fluency literature, yet confers with literature in human computer interaction showing that aesthetics affects impressions of usability (e.g., Kurosu & Kashimura, 1995). Regarding the *intuitive physics* dimension, we found it to be correlated with physical attributes such as 'inelastic v. elastic,' 'liquid v. solid,' and 'light v. heavy.' Thus, research on elasticity (e.g., Paulun et al., 2017), viscosity (e.g., Paulun et al., 2015), and weight (e.g., Schmidt et al., 2020), may be mapping to the same shape information.

One limitation of the current research is that the stimuli were only presented as monoscopically viewed 2D images of 3D objects. The amount of information in 2D images, compared to 3D is limited. Prior research suggests that aesthetics may be affected by the richness of information afforded by the presentation of an object (e.g., Bushong et al., 2010; Gerhard et al., 2016). For example, if an object is viewed as a real 3D object, as opposed to a 2D image of one, their information is processed differently

(Snow et al., 2011). Future research examining the notion of *shapeSpace* should probe its generalizability to impressions of real 3D objects. Furthermore, it would be interesting to also consider the potential neural representation of a *shapeSpace*.

Neighboring research outside of psychology demonstrates the value in developing *shapeSpace.* Human computer interaction researchers have been working on tools to manipulate 3D shapes. In product design, the notion of a *shapeSpace* bears similarity with the notion of shape grammar. The utility of shape grammars is that they afford designers with an efficient ability to explore variations in shape (Alcaide-Marzal et al., 2020). Thus, a psychological model of shape space has the potential to improve computer aided design (CAD) tools. This in turn may lead to improved quality, speed, and a lower cost to product development (S.-W. Hsiao & Chen, 1997).

To conclude, with just a single glance, people extract a wealth of information in their environment (Fei-Fei et al., 2007). They are naturally attuned to statistical regularity in their environment (Zhao et al., 2013). In particular, processing of shape information appears early in development (Emberson et al., 2017). It is possible that people are normally equipped to constantly entrain their individualized *shapeSpace* as a basis for making sense of a world that is filled with novel shapes and to connect their experience with others. *ShapeSpace* may serve as a common ground for sharing experiences between people. The range of shared impressions we documented within the six-dimensional space provides fertile ground for future studies.

5.

Conclusion

Imagine that you went hiking and came across a novel object. Just then, a friend calls you and asks what you are doing. You mention that you are hiking and just came across an object you have not seen before. Next in this moment, could you communicate the novel object's aesthetic?

This dissertation is composed of three projects examining the aesthetics of novel objects. Across the three projects, two different approaches were used. The first project, presented in chapter 2, represents a classic approach to aesthetics—linking an objective property of novel objects to predict preferences. In the following two projects, presented in chapters 3 and 4, I approach aesthetics in terms of shared impressions. Taken together, the research presented in this dissertation suggests that objective properties may be linked to impressions. However, the universality of this link should be interpreted in terms of consensus (i.e., interactionist perspective, see Reber et al., 2004). For example, people often think that symmetry is a universal property of beauty, but recent research

demonstrates that art experts are less likely to agree (Leder et al., 2019). Thus, aesthetics is common sense depending on who you talk to.

In chapter 2, I start by introducing a long history of observations arguing that smoother and curvier shapes are preferred over angular and less curvy shapes. This bias in preferences, otherwise known as the "curvature effect," has been written about for over two centuries. However, one mystery remained unsolved: what is the objective feature that elicits the curvature effect? I hypothesized that 'continuity' is the mathematical definition researchers have been searching for. The results from testing this definition replicates the curvature effect by showing that people prefer greater geometric continuity. This definition was robust across different types of visual stimuli (line drawings, images, animations, virtual, real objects).

The curvature effect is not strong. Its significance is only apparent when comparing it statistically against chance. For industrial designers responsible for producing shapes the majority of people will love, a statistically significant preference is still a valuable one, nonetheless. Yet as a psychologist, it is unsatisfactory to treat aesthetics as a matter of accuracy. For example, what does it mean when a person does not find the curvier object preferable—are they wrong? What seems more appropriate is to focus on aesthetics in terms of shared impressions. Just as how Kant philosophized that aesthetics is as universal as cognition (Hayn-Leichsenring & Chatterjee, 2019), the following two projects approach aesthetics by focusing on shared impressions.

In chapter 3, I tested whether the shape of novel 3-D objects could lead to similar impressions across people. To do this, I introduced a technique from the world of industrial design to psychology called procedural generation and produced stimuli

comprised of shape-families. With these stimuli I explored a few different impressions (approachable, dangerous, beautiful, likable) and examined whether people would have similar impressions of the stimuli because of their shape. Because relatively little is understood regarding how people form impressions of novel objects, I first confirmed the reliability of this behavior by reporting measures of impression consistency and consensus. To situate the magnitude of reliability, I compared the impressions of novel objects to impressions of faces—more familiar and extensively studied stimuli across the current psychological literature.

The results in chapter 3 showed that people always had consistent impressions regardless of whether it was of a face or of an object; and also, regardless of the kind of feature they evaluated. However, in measurements of consensus, impressions of objects appeared more variable than among impressions of faces; varying on the basis of an interaction between features and shape-families. In other words, the findings suggest that aesthetics can be tied to universally objective qualities of visual stimuli, yet how they map together can vary across groups of people. Some impressions (e.g., approachability) may be more reliable across different shape-families than others (e.g., likability).

Finally, in chapter 4, I explored the range of shared impressions of novel objects and discovered a six-dimensional model I refer to as *shapeSpace*. This model efficiently accounts for the majority of variance across shared impressions of novel objects. To facilitate a discussion of each dimension, I provided the following labels: *activity, fluency, intuitive physics, naturalness, clearness, endlessness*. By way of probing the validity of *shapeSpace,* and the meaning represented by its structure*,* the results suggest that its structure may actually shift somewhat from person to person and across different

situations. The first three dimensions possibly shift more often in meaning than the later three. These results suggest the possibility that people's ability to understand the perceived aesthetics of another person is grounded by an entrained *shapeSpace*. Or in Kantian terms, *universal judgments* of *free beauty* may be *common sense* insofar as people share and develop a congruent *shapeSpace.*

References

Aaker, J. L. (1997). Dimensions of brand personality. *Journal of Marketing Research*, *34*(3), 347–356. https://doi.org/10.1177/002224379703400304

Aks, D. J., & Sprott, J. C. (1996). Quantifying aesthetic preference for chaotic patterns. *Empirical Studies of the Arts*, *14*(1), 1–16. https://doi.org/10.2190/6V31-7M9R-T9L5-CDG9

Alcaide-Marzal, J., Diego-Mas, J. A., & Acosta-Zazueta, G. (2020). A 3D shape generative method for aesthetic product design. *Design Studies*, *66*, 144–176. https://doi.org/10.1016/j.destud.2019.11.003

Amir, O., Biederman, I., & Hayworth, K. J. (2011). The neural basis for shape preferences. *Vision Research*, *51*(20), 2198–2206. https://doi.org/10.1016/j.visres.2011.08.015

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, *70*(9), 1–70. https://doi.org/10.1037/h0093718

Banaei, M., Ahmadi, A., Gramann, K., & Hatami, J. (2019). Emotional evaluation of architectural interior forms based on personality differences using virtual reality. *Frontiers of Architectural Research*, 1–10. https://doi.org/10.1016/j.foar.2019.07.005

Bar, M., & Neta, M. (2006). Humans prefer curved visual objects. *Psychological Science*, *17*(8), 645-648. https://doi.org/10.1111/j.1467-9280.2006.01759.x

Bartell, F. O., Dereniak, E. L., & Wolfe, W. L. (1981). The theory and measurement of bidirectional reflectance distribution function (Brdf) and bidirectional transmittance distribution function (BTDF). *Radiation Scattering in Optical Systems*, *0257*, 154–160. https://doi.org/10.1117/12.959611

Berlyne, D. E. (1970). Novelty, complexity, and hedonic value. *Perception & Psychophysics*, *8*(5), 279–286. https://doi.org/10.3758/BF03212593

Bertamini, M, & Palumbo, L. (2015). The aesthetics of curvature in historical context. In *Art and its role in the history of the Balkans (between durability and transientisms)* (Vol. 1–finisher, pp. 729–740). Faculty of Philosophy, Kosovska Mitrovica, University of Pristina.

Bertamini, Marco, Palumbo, L., Gheorghes, T. N., & Galatsidas, M. (2016). Do observers like curvature or do they dislike angularity? *British Journal of Psychology*, *107*(1), 154–178. https://doi.org/10.1111/bjop.12132

Bertamini, Marco, & Sinico, M. (2019). A study of objects with smooth or sharp features created as line drawings by individuals trained in design. *Empirical Studies of the Arts*, 1–17. https://doi.org/10.1177/0276237419897048

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*(2), 115–147. https://doi.org/10.1037/0033-295X.94.2.115

Blazhenkova, O., & Kumar, M. M. (2018). Angular versus curved shapes: Correspondences and emotional processing. *Perception*, *47*(1), 67–89. https://doi.org/10.1177/0301006617731048

Bonnington, C. (2012, November 6). Apple granted patent for original iPad design. *Wired*. https://www.wired.com/2012/11/apple-granted-ipad-patent/

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482–1493. https://doi.org/10.1037/0278-7393.22.6.1482

Burnham, K. P., & Anderson, D. R. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods Research*, *33*, 261–304. https://doi.org/10.1177/0049124104268644

Bushong, B., King, L. M., Camerer, C. F., & Rangel, A. (2010). Pavlovian processes in consumer choice: The physical presence of a good increases willingness-to-pay. *American Economic Review*, *100*(4), 1556–1571. https://doi.org/10.1257/aer.100.4.1556

Chatterjee, A., Widick, P., Sternschein, R., Smith, W. B., & Bromberger, B. (2010). The assessment of art attributes. *Empirical Studies of the Arts*, *28*(2), 207–222. https://doi.org/10.2190/EM.28.2.f

Chemero, A. (2003). An outline of a theory of affordances. *Ecological Psychology*, *15*(2), 181–195. https://doi.org/10.1207/S15326969ECO1502_5

Chen, N., Tanaka, K., Matsuyoshi, D., & Watanabe, K. (2016). Cross preferences for colors and shapes. *Color Research & Application*, *41*(2), 188–195. https://doi.org/10.1002/col.21958

Chen, Y.-C., Huang, P.-C., Woods, A., & Spence, C. (2016). When “Bouba” equals “Kiki”: Cultural commonalities and cultural differences in sound-shape correspondences. *Scientific Reports*, *6*(1), 1–9. https://doi.org/10.1038/srep26681

Chuang, M.-C., & Ma, Y.-C. (2001). Expressing the expected product images in product design of micro-electronic products. *International Journal of Industrial Ergonomics*, *27*(4), 233–245. https://doi.org/10.1016/S0169-8141(00)00053-6

Coburn, A., Vartanian, O., Kenett, Y. N., Nadal, M., Hartung, F., Hayn-Leichsenring, G., Navarrete, G., González-Mora, J. L., & Chatterjee, A. (2020). Psychological and neural responses to architectural interiors. *Cortex*, *126*, 217–241. https://doi.org/10.1016/j.cortex.2020.01.009

Connolly, A. C., Sha, L., Guntupalli, J. S., Oosterhof, N., Halchenko, Y. O., Nastase, S. A., Castello, M. V. di O., Abdi, H., Jobst, B. C., Gobbini, M. I., & Haxby, J. V. (2016). How the human brain represents perceived dangerousness or “predacity” of animals. *Journal of Neuroscience*, *36*(19), 5373–5384. https://doi.org/10.1523/JNEUROSCI.3395-15.2016

Corradi, G., Belman, M., Currò, T., Chuquichambi, E. G., Rey, C., & Nadal, M. (2019). Aesthetic sensitivity to curvature in real objects and abstract designs. *Acta Psychologica*, *197*, 124–130. https://doi.org/10.1016/j.actpsy.2019.05.012

Costa, M., & Bonetti, L. (2016). Geometrical factors in the perception of sacredness. *Perception*, *45*(11), 1240–1266. https://doi.org/10.1177/0301006616654159

Cotter, K. N., Silvia, P. J., Bertamini, M., Palumbo, L., & Vartanian, O. (2017). Curve appeal: Exploring individual differences in preference for curved versus angular

objects. *I-Perception*, *8*(2), 2041669517693023. https://doi.org/10.1177/2041669517693023

Dazkir, S. S., & Read, M. A. (2012). Furniture forms and their influence on our emotional responses toward interior environments. *Environment and Behavior*, *44*(5), 722–732. https://doi.org/10.1177/0013916511402063

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Donderi, D. C. (2006). Visual complexity: A review. *Psychological Bulletin*, *132*(1), 73–97. https://doi.org/10.1037/0033-2909.132.1.73

Duckworth, K. L., Bargh, J. A., Garcia, M., & Chaiken, S. (2002). The automatic evaluation of novel stimuli. *Psychological Science*, *13*(6), 513–519. https://doi.org/10.1111/1467-9280.00490

Echterhoff, G., Higgins, E. T., & Levine, J. M. (2009). Shared reality: Experiencing commonality with others' inner states about the world. *Perspectives on Psychological Science*, *4*(5), 496–521. https://doi.org/10.1111/j.1745-6924.2009.01161.x

Emberson, L. L., Crosswhite, S. L., Richards, J. E., & Aslin, R. N. (2017). The lateral occipital cortex is selective for object shape, not texture/color, at six months. *Journal of Neuroscience*, *37*(13), 3698–3703. https://doi.org/10.1523/JNEUROSCI.3300-16.2017

Engell, A. D., Haxby, J. V., & Todorov, A. (2007). Implicit trustworthiness decisions: Automatic coding of face properties in the human amygdala. *Journal of Cognitive Neuroscience*, *19*(9), 1508–1519. https://doi.org/10.1162/jocn.2007.19.9.1508

Epley, N., Waytz, A., & Cacioppo, J. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

Fantz, R. L., & Miranda, S. B. (1975). Newborn infant attention to form of contour. *Child Development*, *46*(1), 224–228. JSTOR. https://doi.org/10.2307/1128853

Farin, G. (2006). Class A Bézier curves. *Computer Aided Geometric Design*, *23*(7), 573–581.

Farin, G. E., & Farin, G. (2002). *Curves and surfaces for CAGD: A practical guide*. Morgan Kaufmann.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/bf03193146

Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision*, *7*(1), 10–10. https://doi.org/10.1167/7.1.10

Folstein, J. R., Gauthier, I., & Palmeri, T. J. (2012). How category learning affects object representations: Not all morphspaces stretch alike. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(4), 807–820. https://doi.org/10.1037/a0025836

Folstein, J. R., Palmeri, T. J., & Gauthier, I. (2013). Category learning increases discriminability of relevant object dimensions in visual cortex. *Cerebral Cortex*, *23*(4), 814–823. https://doi.org/10.1093/cercor/bhs067

Forsythe, A., Mulhern, G., & Sawey, M. (2008). Confounds in pictorial sets: The role of complexity and familiarity in basic-level picture processing. *Behavior Research Methods*, *40*(1), 116–129. https://doi.org/10.3758/BRM.40.1.116

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *Journal of Neuroscience*, *23*(12), 5235–5246. https://doi.org/10.1523/JNEUROSCI.23-12-05235.2003

Gallant, J. L., Braun, J., & Essen, D. V. (1993). Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science*, *259*(5091), 100–103. https://doi.org/10.1126/science.8418487

Gallant, J. L., Connor, C. E., Rakshit, S., Lewis, J. W., & Van Essen, D. C. (1996). Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. *Journal of Neurophysiology*, *76*(4), 2718–2739. https://doi.org/10.1152/jn.1996.76.4.2718

Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training 'greeble' experts: A framework for studying expert object recognition processes. *Vision Research*, *38*(15), 2401–2428. https://doi.org/10.1016/S0042-6989(97)00442-2

Gerhard, T. M., Culham, J. C., & Schwarzer, G. (2016). Distinct visual processing of real objects and pictures of those objects in 7- to 9-month-old Infants. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00827

Germine, L., Russell, R., Bronstad, P. M., Blokland, G. A. M., Smoller, J. W., Kwok, H., Anthony, S. E., Nakayama, K., Rhodes, G., & Wilmer, J. B. (2015). Individual aesthetic preferences for faces are shaped mostly by environments, not genes. *Current Biology*, *25*(20), 2684–2689. https://doi.org/10.1016/j.cub.2015.08.048

Giovanzana, S. (2011). *A virtual environment for modeling and analysis of human eye* [Ph.D., Università degli Studi di Padova]. http://paduaresearch.cab.unipd.it/3579/

Goldstone, R. L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General*, *130*(1), 116–139. https://doi.org/10.1037/0096-3445.130.1.116

Gómez-Puerto, G., Munar, E., & Nadal, M. (2016). Preference for curvature: A historical and conceptual framework. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00712

Gómez-Puerto, G., Rosselló, J., Corradi, G., Acedo-Carmona, C., Munar, E., & Nadal, M. (2018). Preference for curved contours across cultures. *Psychology of Aesthetics, Creativity, and the Arts*, *12*(4), 432–439. https://doi.org/10.1037/aca0000135

Graham, D., & Field, D. (2008). Statistical regularities of art images and natural scenes: Spectra, sparseness and nonlinearities. *Spatial Vision*, *21*(1–2), 149–164. https://doi.org/10.1163/156856807782753877

Graham, D. J., Friedenberg, J. D., McCandless, C. H., & Rockmore, D. N. (2010). Preference for art: Similarity, statistics, and selling price. *Human Vision and Electronic Imaging XV*, *7527*, 75271A. https://doi.org/10.1117/12.842398

Graham, D. J., & Redies, C. (2010). Statistical regularities in art: Relations with visual coding and perception. *Vision Research*, *50*(16), 1503–1509. https://doi.org/10.1016/j.visres.2010.05.002

Grave, E., Bojanowski, P., Gupta, P., Joulin, A., & Mikolov, T. (2018). Learning word vectors for 157 languages. *ArXiv:1802.06893 [Cs]*. http://arxiv.org/abs/1802.06893

Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition: As soon as you know it is there, you know what it is. *Psychological Science*, *16*(2), 152–160. https://doi.org/10.1111/j.0956-7976.2005.00796.x

Guilford, J. P., & Hackman, R. B. (1936). Varieties and Levels of Clearness Correlated with Eye-Movements. *The American Journal of Psychology*, *48*(3), 371–388. JSTOR. https://doi.org/10.2307/1415883

Guthrie, G., & Wiener, M. (1966). Subliminal perception or perception of partial cue with pictorial stimuli. *Journal of Personality and Social Psychology*, *3*(6), 619–628. https://doi.org/10.1037/h0023197

Hasson, U., Nastase, S. A., & Goldstein, A. (2020). Direct fit to nature: An evolutionary perspective on biological and artificial neural networks. *Neuron*, *105*(3), 416–434. https://doi.org/10.1016/j.neuron.2019.12.002

Hayn-Leichsenring, G. U., & Chatterjee, A. (2019). Colliding terminological systems—Immanuel Kant and contemporary empirical aesthetics. *Empirical Studies of the Arts*, *37*(2), 197–219. https://doi.org/10.1177/0276237418818635

Hekkert, P., Peper, C. (Lieke) E., & Van Wieringen, P. C. W. (1994). The effect of verbal instruction and artistic background on the aesthetic judgment of rectangles.

Empirical Studies of the Arts, *12*(2), 185–203. https://doi.org/10.2190/RMEM-K5KL-R2RN-7A2Y

Hergenhahn, B. R. (2008). *An Introduction to the History of Psychology* (6th ed.). Cengage Learning.

Herzog, A. R., & Bachman, J. G. (1981). Effects of questionnaire length on response quality. *Public Opinion Quarterly*, *45*(4), 549–559. https://doi.org/10.1086/268687

Hevner, K. (1935). Experimental studies of the affective value of colors and lines. *Journal of Applied Psychology*, *19*(4), 385–398. https://doi.org/10.1037/h0055538

Hogarth, W. (1753). *The analysis of beauty: Written with a view of fixing the fluctuating ideas of taste*. Georg Olms Verlag.

Holland, G. (2009, March 2). *A periodic table of form: The secret language of surface and meaning in product design*. Core77. https://www.core77.com/posts/12752/A-Periodic-Table-of-Form-The-secret-language-of-surface-and-meaning-in-product-design-by-Gray-Holland

Holland, M. K., & Wertheimer, M. (1964). Some physiognomic aspects of naming, or, maluma and takete revisited. *Perceptual and Motor Skills*, *19*(1), 111–117. https://doi.org/10.2466/pms.1964.19.1.111

Hönekopp, J. (2006). Once more: Is beauty in the eye of the beholder? Relative contributions of private and shared taste to judgments of facial attractiveness. *Journal of Experimental Psychology. Human Perception and Performance*, *32*(2), 199–209. https://doi.org/10.1037/0096-1523.32.2.199

Hsiao, K., Chen, L., Wang, C., & Tsang, H. (2006). Fundamental dimensions of affective responses to product shapes. *International Journal of Industrial Ergonomics*, 553–564.

Hsiao, S.-W., & Chen, C.-H. (1997). A semantic and shape grammar based approach for product design. *Design Studies*, *18*(3), 275–296. https://doi.org/10.1016/S0142-694X(97)00037-9

Hsu, S. H., Chuang, M. C., & Chang, C. C. (2000). A semantic differential study of designers' and users' product form perception. *International Journal of Industrial Ergonomics*, *25*(4), 375–391. https://doi.org/10.1016/S0169-8141(99)00026-8

Jadva, V., Hines, M., & Golombok, S. (2010). Infants' preferences for toys, colors, and shapes: Sex differences and similarities. *Archives of Sexual Behavior*, *39*(6), 1261–1273. https://doi.org/10.1007/s10508-010-9618-z

Jiang, Y., Gorn, G. J., Galli, M., & Chattopadhyay, A. (2016). Does your company have the right logo? How and why circular- and angular-logo shapes influence brand attribute judgments. *Journal of Consumer Research*, *42*(5), 709–726. https://doi.org/10.1093/jcr/ucv049

Johnson, S. P. (1997). Young infants' perception of object unity: Implications for development of attentional and cognitive skills. *Current Directions in Psychological Science*, *6*(1), 5–11.

Johnson, S. P., Davidow, J., Hall-Haro, C., & Frank, M. C. (2008). Development of perceptual completion originates in information acquisition. *Developmental Psychology*, *44*(5), 1214–1224. https://doi.org/10.1037/a0013215

Jones, K. S. (2003). What is an affordance? *Ecological Psychology*, *15*(2), 107–114. https://doi.org/10.1207/S15326969ECO1502_1

Kahney, L. (2013). *Jony Ive: The genius behind Apple's greatest products*. Penguin.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*(1), 141–151. https://doi.org/10.1177/001316446002000116

Konkle, T., & Oliva, A. (2011). Canonical visual size for real-world objects. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(1), 23–37. https://doi.org/10.1037/a0020413

Kourtzi, Z., & Shiffrar, M. (1999). The visual representation of three-dimensional, rotating objects. *Acta Psychologica*, *102*(2), 265–292. https://doi.org/10.1016/S0001-6918(98)00056-0

Kranjec, A., & Skov, M. (2020). Visualizing aesthetics across two centuries. *Empirical Studies of the Arts*, 0276237420905308. https://doi.org/10.1177/0276237420905308

Krippendorff, K., & Butter, R. (1984). Product semantics: Exploring the symbolic qualities of form. *The Journal of the Industrial Designers Society of America*, *3*(2), 4–9.

Krispel, U., Schinko, C., & Ullrich, T. (2015). A survey of algorithmic shapes. *Remote Sensing*, *7*(10), 12763–12792. https://doi.org/10.3390/rs71012763

Kubricht, J. R., Holyoak, K. J., & Lu, H. (2017). Intuitive physics: Current research and controversies. *Trends in Cognitive Sciences*, *21*(10), 749–759. https://doi.org/10.1016/j.tics.2017.06.002

Kurosu, A. (2019). Design science: An empirical & quantitative approach. *IDSA Education Symposium*, 1–5. https://www.idsa.org/educationpaper/design-science

Kurosu, A., & Todorov, A. (2017). The shape of novel objects contributes to shared impressions. *Journal of Vision*, *17*(13), 1–20. https://doi.org/10.1167/17.13.14

Kurosu, M., & Kashimura, K. (1995). Apparent usability vs. Inherent usability: Experimental analysis on the determinants of the apparent usability. *CHI '95: Conference Companion on Human Factors in Computing Systems*, 292–293. https://doi.org/10.1145/223355.223680

Lebrecht, S., Bar, M., Barrett, L. F., & Tarr, M. J. (2012). Micro-valences: Perceiving affective valence in everyday objects. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00107

Leder, H., & Carbon, C.-C. (2005). Dimensions in appreciation of car interior design. *Applied Cognitive Psychology*, *19*(5), 603–618. https://doi.org/10.1002/acp.1088

Leder, H., Tinio, P. P. L., Brieber, D., Kröner, T., Jacobsen, T., & Rosenberg, R. (2019). Symmetry Is not a universal law of beauty. *Empirical Studies of the Arts*, *37*(1), 104–114. https://doi.org/10.1177/0276237418777941

Lewis, C., & Walker, P. (1989). Typographic influences on reading. *British Journal of Psychology*, *80*(2), 241–257. https://doi.org/10.1111/j.2044-8295.1989.tb02317.x

LoBue, V. (2014). Deconstructing the snake: The relative roles of perception, cognition, and emotion on threat detection. *Emotion*, *14*(4), 701–711. https://doi.org/10.1037/a0035898

Lundholm, H. (1921). The affective tone of lines: Experimental researches. *Psychological Review*, *28*(1), 43–60. https://doi.org/10.1037/h0072647

Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska directed emotional faces—KDEF. In *CD ROM from Department of Clinical Neuroscience, Psychology section*. Karolinska Institutet. https://kdef.se/home/aboutKDEF.html

Magid, E., Soldea, O., & Rivlin, E. (2007). A comparison of Gaussian and mean curvature estimation methods on triangular meshes of range image data. *Computer Vision and Image Understanding*, *107*(3), 139–159. https://doi.org/10.1016/j.cviu.2006.09.007

Marković, S. (2010). Aesthetic experience and the emotional content of paintings. *Psihologija*, *43*(1), 47–64.

Marks, G., & Miller, N. (1987). Ten years of research on the false-consensus effect: An empirical and theoretical review. *Psychological Bulletin*, *102*(1), 72–90. https://doi.org/10.1037/0033-2909.102.1.72

Martinez, J. E., Funk, F., & Todorov, A. (2020). Quantifying idiosyncratic and shared contributions to judgment. *Behavior Research Methods*. https://doi.org/10.3758/s13428-019-01323-0

Mastandrea, S., Zani, A., Giuliani, M. V., & Bove, G. (1992). Meaning of industrial design objects: From designers to users. *Environment and Planning B: Planning and Design*, *19*(3), 307–319. https://doi.org/10.1068/b190307

Matlin, M. W. (1971). Response competition, recognition, and affect. *Journal of Personality and Social Psychology*, *19*(3), 295–300. https://doi.org/10.1037/h0031352

McCrae, R., Zonderman, A., Costa, P., Bond, M., & Paunonen, S. (1996). Evaluating replicability of factors in the revised NEO personality inventory: Confirmatory

factor analysis versus procrustes rotation. *Journal of Personality and Social Psychology*, *70*(3), 552–566.

McManus, I. C. (1980). The aesthetics of simple figures. *British Journal of Psychology*, *71*(4), 505–524. https://doi.org/10.1111/j.2044-8295.1980.tb01763.x

Mickle, T. (2019, June 30). Jony Ive is leaving Apple, but his departure started long ago. *Wall Street Journal*. https://www.wsj.com/articles/jony-ive-is-departing-apple-but-he-started-leaving-years-ago-11561943376

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *ArXiv:1301.3781 [Cs]*. http://arxiv.org/abs/1301.3781

Mothersill, P., & Bove, V. M. (2015). The EmotiveModeler: An emotive form design CAD tool. *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, 339–342. https://doi.org/10.1145/2702613.2725433

Munar, E., Gómez-Puerto, G., Call, J., & Nadal, M. (2015). Common visual preference for curved contours in humans and great apes. *PLoS ONE*, *10*(11), e0141106. https://doi.org/10.1371/journal.pone.0141106

Nagamachi, M. (1995). Kansei engineering: A new ergonomic consumer-oriented technology for product development. *International Journal of Industrial Ergonomics*, *15*(1), 3–11. https://doi.org/10.1016/0169-8141(94)00052-5

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*,

4(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x@10.1111/(ISSN)2041-210X.STATSTOO

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *105*(32), 11087–11092. https://doi.org/10.1073/pnas.0805664105

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. University of Illinois Press.

Palumbo, L., & Bertamini, M. (2016). The curvature effect: A comparison between preference tasks. *Empirical Studies of the Arts*, *34*(1), 35–52. https://doi.org/10.1177/0276237415621185

Palumbo, L., Ruta, N., & Bertamini, M. (2015). Comparing angular and curved shapes in terms of implicit associations and approach/avoidance responses. *PLoS ONE*, *10*(10). https://doi.org/10.1371/journal.pone.0140043

Paulun, V. C., Kawabe, T., Nishida, S., & Fleming, R. W. (2015). Seeing liquids from static snapshots. *Vision Research*, *115*, 163–174. https://doi.org/10.1016/j.visres.2015.01.023

Paulun, V. C., Schmidt, F., Assen, J. J. R. van, & Fleming, R. W. (2017). Shape, motion, and optical cues to stiffness of elastic objects. *Journal of Vision*, *17*(1), 20–20. https://doi.org/10.1167/17.1.20

Peirce, J. W. (2007). PsychoPy-Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. https://doi.org/10.3115/v1/D14-1162

Peytchev, A., & Peytcheva, E. (2017). Reduction of measurement error due to survey length: Evaluation of the split questionnaire design approach. *Survey Research Methods*, *11*(4), 361–368. https://doi.org/10.18148/srm/2017.v11i4.7145

Phillips, F., Norman, J. F., & Beers, A. M. (2010). Fechner's aesthetics revisited. *Seeing and Perceiving*, *23*, 263–271.

Poffenberger, A. T., & Barrows, B. E. (1924). The feeling value of lines. *Journal of Applied Psychology*, *8*(2), 187–205. https://doi.org/10.1037/h0073513

Pourtalebi, S., & Bagherzadeh, H. (2012). Eliciting patterns for product personality design: A pedagogical method for industrial design students. *Procedia - Social and Behavioral Sciences*, *51*, 527–532. https://doi.org/10.1016/j.sbspro.2012.08.201

Provins, K. A., Stockbridge, H. C. W., Forrest, D. W., & Anderson, D. M. (1957). The representation of aircraft by pictorial signs. *Occupational Psychology*, *31*(1), 21–37.

Quinn, P. C., Brown, C. R., & Streppa, M. L. (1997). Perceptual organization of complex visual configurations by young infants. *Infant Behavior and Development*, *20*(1), 35–46. https://doi.org/10.1016/S0163-6383(97)90059-X

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and*

Social Psychology Review, *8*(4), 364–382. https://doi.org/10.1207/s15327957pspr0804_3

Redies, C., Hänisch, J., Blickhan, M., & Denzler, J. (2007). Artists portray human faces with the Fourier statistics of complex natural scenes. *Network: Computation in Neural Systems*, *18*(3), 235–248. https://doi.org/10.1080/09548980701574496

Redies, C., Hasenstein, J., & Denzler, J. (2007). Fractal-like image statistics in visual art: Similarity to natural scenes. *Spatial Vision*, *21*(1–2), 137–148. https://doi.org/10.1163/156856807782753921

Rodway, P., Kirkham, J., Schepman, A., Lambert, J., & Locke, A. (2016). The development of shared liking of representational but not abstract art in primary school children and their justifications for liking. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00021

Russell, P. A. (2000). The aesthetics of rectangle proportion: Effects of judgment scale and context. *The American Journal of Psychology*, *113*(1), 27–42. https://doi.org/10.2307/1423459

Ruta, N., Mastandrea, S., Penacchio, O., Lamaddalena, S., & Bove, G. (2019). A comparison between preference judgments of curvature and sharpness in architectural façades. *Architectural Science Review*, *62*(2), 171–181. https://doi.org/10.1080/00038628.2018.1558393

Ryan, K. J. (2016, August 9). *Samsung says design isn't important. The design world begs to differ*. Inc.Com. https://www.inc.com/kevin-j-ryan/why-the-worlds-top-designers-are-coming-to-defense-of-apple.html

Salgado-Montejo, A., Leon, I. T., Elliot, A. J., Salgado, C. J., & Spence, C. (2015). Smiles over frowns: When curved lines influence product preference. *Psychology & Marketing*, *32*(7), 771–781. https://doi.org/10.1002/mar.20817

Schmidt, F., Fleming, R. W., & Valsecchi, M. (2020). Softness and weight from shape: Material properties inferred from local shape features. *Journal of Vision*, *20*(6), 2–2. https://doi.org/10.1167/jov.20.6.2

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*(7), 671–675. https://doi.org/10.1038/nmeth.2089

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, *4*(8), 299–309. https://doi.org/10.1016/S1364-6613(00)01506-0

Sevenant, M., & Antrop, M. (2010). The use of latent classes to identify individual differences in the importance of landscape dimensions for aesthetic preference. *Land Use Policy*, *27*(3), 827–842. https://doi.org/10.1016/j.landusepol.2009.11.002

Shemesh, A., Talmon, R., Karp, O., Amir, I., Bar, M., & Grobman, Y. J. (2017). Affective response to architecture – investigating human reaction to spaces with different geometry. *Architectural Science Review*, *60*(2), 116–125. https://doi.org/10.1080/00038628.2016.1266597

Sherman, A., Grabowecky, M., & Suzuki, S. (2015). In the working memory of the beholder: Art appreciation is enhanced when visual complexity is compatible with

working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(4), 898–903. https://doi.org/10.1037/a0039314

Shortess, G. K., Clarke, J. C., Richter, M. L., & Seay, M. (2000). Abstract or realistic? Prototypicality of paintings. *Visual Arts Research*, *26*(2), 70–79. JSTOR.

Silvera, D. H., Josephs, R. A., & Giesler, R. B. (2002). Bigger is better: The influence of physical size on aesthetic preference judgments. *Journal of Behavioral Decision Making*, *15*(3), 189–202. https://doi.org/10.1002/bdm.410

Silvia, P. J., & Barona, C. M. (2009). Do people prefer curved objects? Angularity, expertise, and aesthetic preference. *Empirical Studies of the Arts*, *27*(1), 25–42. https://doi.org/10.2190/EM.27.1.b

Snow, J. C., Pettypiece, C. E., McAdam, T. D., McLean, A. D., Stroman, P. W., Goodale, M. A., & Culham, J. C. (2011). Bringing the real world into the fMRI scanner: Repetition effects for pictures versus real objects. *Scientific Reports*, *1*(1), 130. https://doi.org/10.1038/srep00130

So, C. (2019). What Makes Good Design? Revealing the Predictive Power of Emotions and Design Dimensions in Non-Expert Design Vocabulary. *The Design Journal*, *22*(3), 325–349. https://doi.org/10.1080/14606925.2019.1589204

Soranzo, A., Petrelli, D., Ciolfi, L., & Reidy, J. (2018). On the perceptual aesthetics of interactive objects. *Quarterly Journal of Experimental Psychology*, *71*(12), 2586–2602. https://doi.org/10.1177/1747021817749228

Spehar, B., Clifford, C. W. G., Newell, B. R., & Taylor, R. P. (2003a). Universal aesthetic of fractals. *Computers and Graphics*, *27*(5), 813–820. https://doi.org/10.1016/S0097-8493(03)00154-7

Spehar, B., Clifford, C. W. G., Newell, B. R., & Taylor, R. P. (2003b). Universal aesthetic of fractals. *Computers & Graphics*, *27*(5), 813–820. https://doi.org/10.1016/S0097-8493(03)00154-7

Summit, J., & Sommer, R. (1999). Further studies of preferred tree shapes. *Environment and Behavior*, *31*(4), 550–576. https://doi.org/10.1177/00139169921972236

Sutherland, C. A. M., Oldmeadow, J. A., Santos, I. M., Towler, J., Michael Burt, D., & Young, A. W. (2013). Social inferences from faces: Ambient images generate a three-dimensional model. *Cognition*, *127*(1), 105–118. https://doi.org/10.1016/j.cognition.2012.12.001

Swami, V., & Furnham, A. (2012). The effects of symmetry and personality on aesthetic preferences. *Imagination, Cognition and Personality*, *32*(1), 41–57. https://doi.org/10.2190/IC.32.1.d

Todorov, A., & Duchaine, B. (2008). Reading trustworthiness in faces without recognizing faces. *Cognitive Neuropsychology*, *25*(3), 395–410. https://doi.org/10.1080/02643290802044996

Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology*, *66*(1), 519–545. https://doi.org/10.1146/annurev-psych-113011-143831

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, *12*(12), 455–460. https://doi.org/10.1016/j.tics.2008.10.001

Treisman, A. (1986). Features and objects in visual processing. *Scientific American*, *255*(5), 114B – 125. JSTOR.

Uher, J. (1991). On zigzag designs: Three levels of meaning. *Current Anthropology*, *32*(4), 437–439. https://doi.org/10.1086/203979

Van Geert, E., & Wagemans, J. (2020). Order, complexity, and aesthetic appreciation. *Psychology of Aesthetics, Creativity, and the Arts*, *14*(2), 135–154. https://doi.org/10.1037/aca0000224

van Oel, C. J., & van den Berkhof, F. W. (Derk). (2013). Consumer preferences in the design of airport passenger areas. *Journal of Environmental Psychology*, *36*, 280–290. https://doi.org/10.1016/j.jenvp.2013.08.005

Vartanian, O., Navarrete, G., Chatterjee, A., Fich, L. B., Leder, H., Modroño, C., Nadal, M., Rostrup, N., & Skov, M. (2013). Impact of contour on aesthetic judgments and approach-avoidance decisions in architecture. *Proceedings of the National Academy of Sciences*, *110*, 10446–10453. https://doi.org/10.1073/pnas.1301227110

Vartanian, O., Navarrete, G., Chatterjee, A., Fich, L. B., Leder, H., Modroño, C., Rostrup, N., Skov, M., Corradi, G., & Nadal, M. (2019). Preference for curvilinear contour in interior architectural spaces: Evidence from experts and nonexperts. *Psychology of Aesthetics, Creativity, and the Arts*, *13*(1), 110–116. https://doi.org/10.1037/aca0000150

Vessel, E. A., & Rubin, N. (2010). Beauty and the beholder: Highly individual taste for abstract, but not real-world images. *Journal of Vision*, *10*(2), 1–14. https://doi.org/10.1167/10.2.18

Wang, S., Lilienfeld, S. O., & Rochat, P. (2015). The uncanny valley: Existence and explanations. *Review of General Psychology*, *19*(4), 393–407. https://doi.org/10.1037/gpr0000056

Westerman, S. J., Gardner, P. H., Sutherland, E. J., White, T., Jordan, K., Watts, D., & Wells, S. (2012). Product design: Preference for rounded versus angular design elements. *Psychology & Marketing*, *29*(8), 595–605. https://doi.org/10.1002/mar.20546

Whittlesea, B. W. A. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(6), 1235–1253. https://doi.org/10.1037/0278-7393.19.6.1235

Wilcox, T. (1999). Object individuation: Infants' use of shape, size, pattern, and color. *Cognition*, *72*(2), 125–166. https://doi.org/10.1016/S0010-0277(99)00035-9

Wilkinson, F., Wilson, H. R., & Habak, C. (1998). Detection and recognition of radial frequency patterns. *Vision Research*, *38*(22), 3555–3568. https://doi.org/10.1016/S0042-6989(98)00039-X

Williams, P., & Simons, D. J. (2000). Detecting changes in novel, complex three-dimensional objects. *Visual Cognition*, *7*(1–3), 297–322. https://doi.org/10.1080/135062800394829

Wingfield, N. (2012, August 24). Jury awards $1 billion to Apple in Samsung patent case. *The New York Times*. https://www.nytimes.com/2012/08/25/technology/jury-reaches-decision-in-apple-samsung-patent-trial.html

Wronecki, J. (2007). Surface modeling techniques for automotive and product design. *2007 Annual Conference & Exposition*, *1163*, 1–7. https://peer.asee.org/surface-modeling-techniques-for-automotive-and-product-design

Yamane, Y., Carlson, E. T., Bowman, K. C., Wang, Z., & Connor, C. E. (2008). A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nature Neuroscience*, *11*(11), 1352–1360. https://doi.org/10.1038/nn.2202

Yang, C.-C., & Chang, H.-C. (2012). Selecting representative affective dimensions using Procrustes analysis: An application to mobile phone design. *Applied Ergonomics*, *43*(6), 1072–1080. https://doi.org/10.1016/j.apergo.2012.03.008

Yumer, M. E., Chaudhuri, S., Hodgins, J. K., & Kara, L. B. (2015). Semantic shape editing using deformation handles. *ACM Transactions on Graphics*, *34*(4), 86:1–86:12. https://doi.org/10.1145/2766908

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, *35*(2), 151–175. https://doi.org/10.1037/0003-066X.35.2.151

Zajonc, Robert B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2, Pt.2), 1–27. https://doi.org/10.1037/h0025848

Zhao, J., Al-Aidroos, N., & Turk-Browne, N. B. (2013). Attention Is spontaneously biased toward regularities. *Psychological Science*, *24*(5), 667–677. https://doi.org/10.1177/0956797612460407

Appendix

6.1 Chapter Appendix

6.1.1 Study-1 Appendix

6.1.1.1 Consensus methods

Table 15 shows the results of various tests for significance using two methods for calculating consensus. None of the analyses were contradictory, however the estimated effect size differed between methods for calculating consensus.

Table 15. Comparing methods for analyzing consensus.

(Study-1) Two consensus methods

	Blocked Pairwise Correlation Correlations amongst all possible pairing of individuals per exposure.	**Correlation with the Mean** Correlations between individuals and the average reflecting all others.
Significance across all evaluations of objects	$t(39) = 6.67$, $p < .001$, $d = 1.06$, $r_M = .33$, $r_{SD} = .3$	$t(99) = 10.48$, $p < .001$, $d = 1.05$, $r_M = .55$, $r_{SD} = .52$
Significance across all evaluations of faces	$t(39) = 32.83$, $p < .001$, $d = 5.19$, $r_M = .26$, $r_{SD} = .05$	$t(87) = 27.46$, $p < .001$, $d = 2.93$, $r_M = .63$, $r_{SD} = .17$
Interaction effect between stimuli and evaluations	$F(3, 72) = 1003.4$, $p < .001$, $\eta_p^2 = .98$, $\eta^2 = .55$	$F(3, 180) = 40.69$, $p < .001$, $\eta_p^2 = .40$, $\eta^2 = .28$
Main effect of stimuli (between objects and faces)	$F(1, 72) = 272.8$, $p < .001$, $\eta_p^2 = .79$, $\eta^2 = .05$	$F(1, 180) = 12.80$, $p < .001$, $\eta_p^2 = .07$, $\eta^2 = .03$
Main effect of evaluation	$F(3, 72) = 716.8$, $p < .001$, $\eta_p^2 = .97$, $\eta^2 = .39$	$F(3, 180) = 40.42$, $p < .001$, $\eta_p^2 = .41$, $\eta^2 = .29$
Between exposures for objects	$F(9, 30) = .01$, $p = 1$, $\eta_p^2 = .00$, $\eta^2 = .00$	n/a
Between exposures for faces	$F(9, 30) = .20$, $p = .99$, $\eta_p^2 = .06$, $\eta^2 = .06$	n/a

Notes. The cells marked 'n/a' were not possible because averages across blocks were first calculated. To test for the significance amongst the pairwise correlation, the analysis was calculated for each exposure rather than by using the averages across all the exposure blocks.

6.1.1.2 Two vs. ten blocks

Within the studies, significance testing used Fisher's *r* to *z* transformed scores. For the plots, the *z* scores were transformed back to r coefficients. The transformations appear to inflate the correlation value; therefore, we report the correlation values at each stage of its transformation. We also report an analysis of impressions across 10 exposures and two exposures. See **Table 16** to examine the effects on consistency, and

Table 17 for the effects on consensus. It is debatable whether averages represent truer values because they decrease noise, or if they overinflate correlations while failing to represent any actual data point; therefore, we present both values.

Table 16. Consistency for each condition; across 10 blocks was similar to across the first 2 blocks.

(Study-1) Consistency: Faces & Novel Objects

	Pearson's r		Fisher's r to z		Fisher's z to r	
	Exposures 1-10	Exposures 1-2	Exposures 1-10	Exposures 1-2	Exposures 1-10	Exposures 1-2
Approach & Face	r_M = .60 r_{SD} = .19	r_M = .51 r_{SD} = .18	z_M = .74 z_{SD} = .29	z_M = .59 z_{SD} = .25	r_M = .63 r_{SD} = .29	r_M = .53 r_{SD} = .24
Danger & Face	r_M = .47 r_{SD} = .24	r_M = .45 r_{SD} = .22	z_M = .56 z_{SD} = .34	z_M = .52 z_{SD} = .28	r_M = .51 r_{SD} = .33	r_M = .48 r_{SD} = .27
Beauty & Face	r_M = .59 r_{SD} = .21	r_M = .54 r_{SD} = .21	z_M = .74 z_{SD} = .33	z_M = .65 z_{SD} = .30	r_M = .63 r_{SD} = .32	r_M = .57 r_{SD} = .29
Like & Face	r_M = .59 r_{SD} = .23	r_M = .55 r_{SD} = .20	z_M = .73 z_{SD} = .32	z_M = .65 z_{SD} = .26	r_M = .62 r_{SD} = .31	r_M = .57 r_{SD} = .26
Approach & Object	r_M = .72 rSD = .21	r_M = .75 r_{SD} = .10	z_M = 1.01 z_{SD} = .38	z_M = 1.02 z_{SD} = .22	r_M = .77 r_{SD} = .37	r_M = .77 r_{SD} = .21
Danger & Object	r_M = .77 r_{SD} = .11	r_M = .77 r_{SD} = .12	z_M = 1.08 z_{SD} = .27	z_M = 1.09 z_{SD} = .28	r_M = .79 r_{SD} = .26	r_M = .80 r_{SD} = .27
Beauty & Object	r_M = .58 r_{SD} = .22	r_M = .57 r_{SD} = .22	z_M = .72 zSD = .33	z_M = .71 z_{SD} = .36	r_M = .62 r_{SD} = .32	r_M = .61 r_{SD} = .34

Like & Object	r_M = .58 r_{SD} = .21	r_M = .55 r_{SD} = .21	z_M = .72 z_{SD} = .34	z_M = .67 z_{SD} = .32	r_M = .62 r_{SD} = .33	r_M = .59 r_{SD} = .31

Notes. Transformations from r to z occurred before averaging.

Table 17. Consensus for each condition; across 10 blocks was similar to across the first 2 blocks.

(Study-1) Consensus: Faces & Novel Objects

	Pearson's r		Fisher's r to z		Fisher's z to r	
	Exposures 1-10	Exposures 1-2	Exposures 1-10	Exposures 1-2	Exposures 1-10	Exposures 1-2
Approach & Face	r_M = .65 r_{SD} = .15	r_M = .57 r_{SD} = .19	z_M = .81 z_{SD} = .23	z_M = .68 z_{SD} = .26	r_M = .67 r_{SD} = .23	r_M = .59 r_{SD} = .25
Danger & Face	r_M = .58 r_{SD} = .22	r_M = .52 r_{SD} = .20	z_M = .72 z_{SD} = .32	z_M = .62 z_{SD} = .27	r_M = .62 r_{SD} = .31	r_M = .55 r_{SD} = .27
Beauty & Face	r_M = .68 r_{SD} = .13	r_M = .61 r_{SD} = .15	z_M = .87 z_{SD} = .23	z_M = .73 z_{SD} = .21	r_M = .70 r_{SD} = .23	r_M = .62 r_{SD} = .21
Like & Face	r_M = .60 r_{SD} = .17	r_M = .54 r_{SD} = .17	z_M = .74 z_{SD} = .28	z_M = .63 z_{SD} = .23	r_M = .63 r_{SD} = .27	r_M = .56 r_{SD} = .23
Approach & Object	r_M = .85 r_{SD} = .35	r_M = .72 r_{SD} = .47	z_M = 1.62 z_{SD} = .69	z_M = 1.16 z_{SD} = .83	r_M = .92 r_{SD} = .60	r_M = .82 r_{SD} = .68
Danger & Object	r_M = .94 r_{SD} = .04	r_M = .89 r_{SD} = .05	z_M = 1.83 z_{SD} = .32	z_M = 1.49 z_{SD} = .26	r_M = .95 r_{SD} = .31	r_M = .90 r_{SD} = .26
Beauty & Object	r_M = .03 r_{SD} = .20	r_M = .05 r_{SD} = .29	z_M = .04 z_{SD} = .21	z_M = .05 z_{SD} = .30	r_M = .04 r_{SD} = .21	r_M = .05 r_{SD} = .29
Like & Object	r_M = .38 r_{SD} = .62	r_M = .34 r_{SD} = .56	z_M = .52 z_{SD} = .91	z_M = .43 z_{SD} = .73	r_M = .48 r_{SD} = .72	r_M = .40 r_{SD} = .62

Notes. Transformations from r to z occurred before averaging.

6.1.1.3 Comparing consensus between conditions

Table 18 compares the level of consensus between each condition. On the left, three methods are reported: pairwise correlations computed upon each block ("Blocked Pairwise Correlation"), pairwise correlation using the averaged ratings across the ten blocks ("Pairwise Correlation"), and correlations with the mean. On the right, we compare each condition to test the significance of their difference using the values

generated from the "correlation with the mean" method. The arrows point towards the condition yielding significantly more consensus.

Table 18. Comparing consensus across conditions.

(Study-1) Comparing consensus

Blocked Pairwise Correlation	Pairwise Correlation	Correlation with the Mean		Approach & Face	Danger & Face	Beauty & Face	Like & Face	Approach & Object	Danger & Object	Beauty & Object
$r_M = .29$ $r_{SD} = .01$	$r_M = .44$	$r_M = .65$ $r_{SD} = .15$	Approach & Face							
$r_M = .20$ $r_{SD} = .03$	$r_M = .36$	$r_M = .58$ $r_{SD} = .22$	Danger & Face	• $p = 1$						
$r_M = .31$ $r_{SD} = .02$	$r_M = .49$	$r_M = .68$ $r_{SD} = .13$	Beauty & Face	• $p = 1$	• $p = 1$					
$r_M = .25$ $r_{SD} = .01$	$r_M = .39$	$r_M = .60$ $r_{SD} = .17$	Like & Face	• $p = 1$	• $p = 1$	• $p = 1$				
$r_M = .54$ $r_{SD} = .05$	$r_M = .72$	$r_M = .85$ $r_{SD} = .35$	Approach & Object	← $p < .001$	← $p < .001$	← $p < .001$	← $p < .001$			
$r_M = .70$ $r_{SD} = .02$	$r_M = .89$	$r_M = .94$ $r_{SD} = .04$	Danger & Object	← $p < .001$	← $p < .001$	← $p < .001$	← $p < .001$	• $p = 1$		
$r_M = .00$ $r_{SD} = .00$	$r_M = .00$	$r_M = .03$ $r_{SD} = .20$	Beauty & Object	↑ $p < .001$	↑ $p < .001$	↑ $p < .001$	↑ $p < .001$	↑ $p < .001$	↑ $p < .001$	
$r_M = .08$ $r_{SD} = .02$	$r_M = .15$	$r_M = .17$ $r_{SD} = .62$	Like & Object	• $p = .31$	• $p = 1$	• $p = .18$	• $p = 1$	↑ $p < .001$	↑ $p < .001$	← $p = .006$

Notes. Significant testing was executed using Fisher's r to z transformations on the correlations with the mean method; the other method does not allow for this level of analysis. If there was a significant difference, an arrow marks the variable associated with the greater consensus. The standard deviation for the blocked pairwise consensus represents differences in consensus across ten blocks. Whereas pairwise correlations averages across the ten blocks before calculating pairwise correlations; this later method does not provide an interpretable standard deviation statistic. The standard deviation for the correlation with the mean represents the differences between participants using impressions averaged over ten blocks.

6.1.1.3 Predicting impressions

Here we provide the regression tables for the models in which shapes predict impressions, see **Table 19** for the results of study-1. Separate mixed-effects models were made for each condition which yielded a significant level of consensus.

Table 19. Regressing impressions on shape-parameters.

(Study-1) Regression per shape-family

		N	R^2	β	SE	t	SD
Approach		25	.74				
	Fixed Effects		.54				
	Intercept			4.56	.18	25.48	
	Points			-.55***	.03	20.95	
	Flux			-1.17***	.03	43.13	
	Points:Flux			-.24***	.03	9.01	
	Random Effects		.20				
	Participants						.89
	Residual						1.04
Danger		25	.86				
	Fixed Effects		.66				
	Intercept			4.96	.20	24.54	
	Points			.41***	.02	18.82	
	Flux			1.60***	.02	71.73	
	Points:Flux			.29***	.02	12.93	
	Random Effects		.20				
	Participants						1.00
	Residual						.85
Like		25	.18				
	Fixed Effects		.10				
	Intercept			4.79	.11	43.49	
	Points			-.08*	.04	2.05	
	Flux			-.46***	.04	11.10	
	Points:Flux			-.16***	.04	3.89	
	Random Effects		.08				
	Participants						.51
	Residual						1.58

*Notes. b = Raw regression coefficients coefficient. SE = Standard Error for β. ' : ' denotes an interaction effect. Analysis using R packages: lme4 (1.1-12), lmerTest (2.0-32), & MuMIn (1.15.6). ***p < .001, *p < .05*

6.1.2 Study-2 Appendix

6.1.2.1 Consensus methods

Table 20 shows the level of consensus from different methods of calculation and across transformations from Pearson's *r* coefficients to Fisher's *z* and back to Pearson's *r*. Again, three measures of consensus are reported: pairwise correlations computed upon each block ("Blocked Pairwise Correlation"), pairwise correlation using the averaged ratings across the two blocks ("Pairwise Correlation"), and correlations with the mean. Averages are reported. For the *r* to *z* values, averaging was done after this transformation.

The resulting *z* scores were then transformed back into *r* coefficients. These transformed scores are only possible using the "Blocked Pairwise Correlation" method and the "Correlation with the Mean" method because they generate more than value per condition. **Table 21** compares the level of consensus between conditions.

Table 20. Comparing methods for analyzing consensus.

(Study-2) Two consensus methods

	Blocked Pairwise Correlation				Pairwise Correlation		Correlation with the Mean		
	**N*	*r*	*r→ z*	*r→z→r*	*N*	*r*	*N*	*r*	*r→z*	*r→z→r*
Danger										
Family-X	14	.68	.82	.68	18	.80	18	.89	1.57	.92
Family-Y		.17	.17	.17		.30		.53	.74	.63
Family-Z		.13	.13	.13		.11		.31	.45	.43
Beauty										
Family-X	19	.00	.00	.00	20	.03	20	-.01	-.02	-.02
Family-Y		.30	.31	.30		.38		.60	.83	.68
Family-Z		.16	.16	.16		.19		.42	.57	.52

*Notes. Data regarding the images generated using the new renderer; 30 per shape-family. Means are reported. *N for Blocked Pairwise Correlation excludes four participants from the dangerous condition and one participant's data from the beautiful condition due to an absence of observed variance across their impressions from a single block. The block which lacked variance occurred twice as the first block and three times as the second block amongst these five excluded participants.*

Table 21. Comparing consensus across conditions.

(Study-2) Consensus compared across conditions

	Danger & *Family-X*	Danger & *Family-Y*	Danger & *Family-Z*	Beauty & *Family-X*	Beauty & *Family-Y*
Danger & *Family-Y*	↑ $p < .001$				
Danger & *Family-Z*	↑ $p < .001$	• $p = .57$			
Beauty & *Family-X*	↑ $p < .001$	↑ $p < .001$	• $p = .05$		
Beauty & *Family-Y*	↑ $p < .001$	• $p = 1$	• $p = .20$	← $p < .001$	
Beauty & *Family-Z*	↑ $p < .001$	• $p = 1$	• $p = 1$	← $p < .01$	• $p = .57$

Notes. Significant testing was executed using Fisher's r to z transformations on the correlations with the mean method; the other method does not allow for this level of analysis. If there was a significant difference, an arrow marks the variable associated with the greater consensus. Comparisons between different evaluations of the same shape-family are highlighted in black.

6.1.2.2 Predicting impressions

Here we provide the regression tables for the models in which shapes predict impressions; **Table 22** shows the results of study-2. Separate mixed-effects models were made for each condition which yielded a significant level of consensus.

Table 22. Regressing impressions on shape-parameters.

(Study-2) Regression per shape-family

	N	R^2	*β*	*SE*	*t*	*SD*
Danger & Family-X	18	.80				
Fixed Effects		.53				
Intercept			5.00	.30	16.42	
Points			.51***	.05	10.26	
Flux			1.57***	.05	30.26	
Points:Flux			.13***	.05	2.53	
Random Effects		.27				
Participants						1.27
Residual						1.09
Danger & Family-Y	18	.61				
Fixed Effects		.15				
Intercept			2.34	.25	9.34	
Points			-.18*	.08	-2.33	
Flux			-.18*	.08	-2.18	
Radius			-.35***	.07	-4.71	
Length			.03	.08	.33	
Points:Flux			.01	.08	.08	
Points:Radius			-.08	.11	-.71	
Flux:Radius			.38***	.10	3.74	
Points:Length			-.22*	.10	-2.19	
Flux:Length			.44***	.13	3.42	
Radius:Length			-.46***	.09	-5.31	
Points:Flux:Radius			.24	.12	1.96	
Points:Flux:Length			-.05	.13	-.42	
Points:Radius:Length			-.38*	.15	-2.50	
Flux:Radius:Length			.06	.11	.53	
Points:Flux:Radius:Length			.15	.14	1.01	
Random Effects		.46				
Participants						1.02
Residual						.95
Danger & Family-Z	18	.50				
Fixed Effects		.10				
Intercept			1.88	.18	10.66	

Points			.25***	.04	6.51	
Flux			.17***	.04	4.43	
Points:Flux			.11**	.04	2.93	
Random Effects		.40				
Participants						.73
Residual						.83
Beauty & Family-Y	20	.61				
Fixed Effects		.23				
Intercept			6.54	.29	22.44	
Points			.41***	.10	4.30	
Flux			.34**	.10	3.28	
Radius			.45***	.10	4.90	
Length			.00	.10	.05	
Points:Flux			-.01	.10	-.08	
Points:Radius			-.14	.14	-1.03	
Flux:Radius			-.49***	.13	-3.90	
Points:Length			.46***	.12	3.71	
Flux:Length			-.73***	.16	-4.52	
Radius:Length			.45***	.11	4.22	
Points:Flux:Radius			- .45**	.15	-2.95	
Points:Flux:Length			.10	.16	.63	
Points:Radius:Length			.94***	.19	5.00	
Flux:Radius:Length			.05	.14	.37	
Points:Flux:Radius:Length			-.27	.18	-1.49	
Random Effects		.38				
Participants						1.24
Residual						1.24
Beauty & Family-Z	20	.68				
Fixed Effects		.09				
Intercept			3.39	.36	9.33	
Points			-.43***	.05	8.42	
Flux			-.34***	.05	6.42	
Points:Flux			-.04	.05	.73	
Random Effects		.59				
Participants						1.61
Residual						1.17

*Notes. Fixed Effects = shape-parameters. R^2 calculated using the method by Nakagawa & Schielzeth (2013). b = Raw regression coefficients coefficient. SE = Standard Error for b. ':' denotes an interaction effect. Analysis using R packages: lme4 (1.1-12), lmerTest (2.0-32), & MuMIn (1.15.6). ***p < .001, **p < .01, *p < .05*

Future efforts aiming to understand shape effects may seek to extrapolate the shape-parameters presented here. For such endeavors, we provide a few suggestions. First, the shape-parameters tested here should be examined within the context of its algorithm. For example, in **Table 22** we can see points flipping between a positive and negative β value across conditions. Also, the labels we use to describe each shape-parameter have been symbolically assigned and could therefore be interpreted and

recreated in a way which results in dissimilarities. For example: *family-y's radius* could be described as the radius of distally extruding volumes, rather than the radius of the entire object. Second, not all shape-parameters may be equally discriminable (Folstein et al., 2012). Not knowing the discriminability of each shape-parameter complicates generalizations of their power across different algorithms. Finally, multicollinearity was present amongst our randomly generated values. This further complicates the inferences that can be made regarding the relative contribution of individual shape-parameters. For example: the *length* shape-parameter is significant when the *flux* shape-parameter is not included in *family-y's* models. Therefore, as an example, it would be premature to claim that more extreme *flux* for *any* object would lead to heightened impressions of dangerousness. To investigate specific metrics, we suggest that future research document the discriminability of each shape-parameter and samples values for them non-randomly.

6.1.3 Comparing Alternative Models

6.1.3.1 Alternative variables

Could the effect shapes have on impressions be explained by other perceptual variables? Prior research has examined many specific hypotheses regarding how properties of visual stimuli relate to preferences, such as familiarity (e.g., Reber et al., 2004; Zajonc, 1980), an image's spatial frequency amplitude spectrum slope (e.g., Graham et al., 2010; Graham & Redies, 2010), fractal dimension (e.g., Aks & Sprott, 1996; Spehar et al., 2003), complexity (e.g., Phillips et al., 2010; Shortess et al., 2000), size (e.g., Chen et al., 2016; Konkle & Oliva, 2011), and curvature (e.g., Bar & Neta, 2006; Bertamini et al., 2016; Vartanian et al., 2013). Our stimuli, however, were designed

to assess the reliability of shape's effects in general. Therefore, *a priori* the shape-parameters we used were not normalized according to previously explored perceptual variables (i.e., how objects belonging to *family-x* could map to metrics of curviness is not readily apparent). Nevertheless, it is still possible to measure the stimuli in terms of these perceptual variables. Thus, here we share the results of testing alternative perceptual variables to see if they already explain shapes effect on impressions of novel objects. The same 90 novel objects from study-2 were measured in terms of seven perceptual variables: subjective familiarity ('familiarity'), spatial frequency amplitude spectrum slope ('amplitude spectrum slope'), fractal dimension ('fractility'), image complexity ('complexity'), surface area ('area'), object volume ('volume'), and surface curvature ('curvature'). The stimuli consist of 30 objects generated from three shape-families (*family-x, family-y,* & *family-z*). Three of the perceptual variables measure the 3D object (volume, area, curvature), while the other four perceptual variables measure the 2D images (i.e., "new renderings").

6.1.3.2 Shape-parameters' correlation to other variables

Each of the three shape-families (*family-x, family-y, family-z*) were measured in terms of seven variables: familiarity, amplitude, fractality, complexity, area, volume, and curvature. Each of these perceptual variables was correlated with the shape-parameters, but there was no apparent regularity. The Pearson's *r* correlation between each variable and shape-parameter are shown in **Table 23**. Although differences between each object were dependent upon changes in the value of a shape-parameter, none of the perceptual

variables were perfectly correlated with the shape-parameters. Thus, none of these perceptual variables alone explains the specific effects of shapes on shared impressions.

Table 23. Relating shape-parameters with other visual statistics.

orrelation between shape-parameters, 2D image statistics, and 3D geometry statistics.

	Familiarity	Amplitude	Fractality	Complexity	Area	Volume	Curvature		
							Gaussian Max	Gaussian Min	Mean Max
Family-X									
Points	.48*	.55**	-.32	.67***	.23	.14	.39*	-.32	.48**
Flux	.91***	.73***	-.91***	.82***	-.76***	-.86**	.50**	.16	.52**
Family-Y									
Points	.72***	.53**	.20	.83***	.17	.27	-.16	*NA*	-.13
Flux	.16	-.07	-.09	.10	-.29	-.26	-.07	*NA*	.19
Radius	.02	-.28	.48**	.11	-.46*	-.18	.30	*NA*	-.57**
Length	-.36*	-.23	-.41*	-.37*	-.26	-.64***	.07	*NA*	.46*
Family-Z									
Points	-.31	.52**	.34	.86***	74***	77***	-.06	-.28	-.07
Flux	-.59***	.09	.26	.51**	69***	.30	-.23	.26	-.24

*Notes. The image sizes were recorded in bytes. The original renders were 1000x1000 pixels. NA: all objects from family-y had a Gaussian-min value of 0 and therefore correlations could not be calculated. *p<.05 **p<.01 ***p<.001*

6.1.3.3 Comparing models predicting impressions

The inference goal for comparing models of impressions is to determine if the models of impressions using a perceptual variable informed by prior research supersedes the model of impressions using shape-parameters in goodness of fit. Each model was a mixed effects linear regression with participants set as a random effect and a perceptual variable (or the shape-parameters) set as the fixed effects. We compare the goodness of each model's fit using a second-order Akaike Information Criterion (AIC_c) (Burnham & Anderson, 2004). The perceptual variables were analyzed as discrete models because of multicollinearity. As shown in Table 24, many of the variables and shape-parameters are indeed highly correlated—an unsurprising fact given that the perceptual variables are the

consequent of having manipulated the shape-parameters. We hypothesized that the models of impressions using shapes would consistently feature the lowest AIC_c scores to show that the effects of shapes cannot be readily explained by previously explored perceptual variables. In addition to the AIC_c scores, we also report the proportion of variance, R^2, explained by the fixed effects within each model.

Table 24. Models of object impressions based on shape-parameters often outperform models based on other statistics. Each model is ranked by goodness of fit (lowest AIC_c score). Variance explained by the fixed effects (R^2_{fixed}) and total model (R^2_{total}) are also included.

Comparing models of impressions based on shapes to others.

Variable AIC_c R^2_{fixed} R^2_{total}	Danger			Beauty		
	Family-X	*Family-Y*	*Family-Z*	*Family-X*	*Family-Y*	*Family-Z*
1	Shape 1705.99 .53 .80	Shape 1573.85 .15 .61	Area 1390.23 .10 .50	Fractality 2269.83 .00 .36	Shape 2069.14 .23 .61	Area 1967.90 .09 .69
2	Familiarity 1723.09 .53 .79	Fractality 1550.21 .15 .60	Shape 1395.50 .10 .50	Complexity 2270.36 .00 .36	Fractality 2136.83 .16 .54	Shape 1983.85 .09 .68
3	Complexity 1804.47 .49 .76	Curvature 1573.10 .13 .59	Complexity 1406.16 .09 .48	Area 2270.50 .00 .36	Curvature 2166.87 .14 .52	Complexity 1990.69 .08 .68
4	Fractality 1923.09 .43 .70	Volume 1680.13 .04 .49	Volume 1427.88 .07 .46	Volume 2270.67 .00 .36	Volume 2261.21 .06 .43	Volume 2007.33 .07 .67
5	Amplitude 2027.57 .37 .63	Area 1699.97 .02 .47	Familiarity 1465.54 .03 .42	Familiarity 2271.59 .00 .36	Complexity 2268.87 .05 .43	Familiarity 2081.08 .03 .62
6	Volume 2176.60 .25 .51	Complexity 1711.15 .01 .46	Amplitude 1469.62 .02 .42	Amplitude 2271.67 .00 .36	Familiarity 2278.55 .04 .42	Amplitude 2093.31 .02 .62
7	Curvature 2239.80 .19 .45	Familiarity 1714.17 .01 .46	Curvature 1471.54 .03 .42	Curvature 2273.51 .00 .36	Amplitude 2306.86 .01 .39	Fractality 2098.30 .02 .62
8	Area 2275.01 .15 .41	Amplitude 1721.60 .01 .45	Fractality 1480.86 .01 .40	Shape 2273.73 .00 .36	Area 2309.94 .01 .38	Curvature 2102.10 .02 .61

Notes. Impression data (from study-2; danger, N = 18; beauty, N = 20). Beauty of Family-x yielded no consensus so modeling the group results in 0 variance explained by fixed effects. Yet, they are still included in gray for continuity. Models were analyzed using R package lme4 (version 1.12-12) and AICc scores were calculated using the R package MuMIm (version 1.15.6). Separate models entered in the following manner: lmer(impressions ~ variable + (1 | participant). 'Impressions' is the average impression across each participant's two evaluations. 'Variable' is a statistic, such as shapes-parameter values. Models based on shape-parameters are highlighted in black. All variables were scaled and centered per family. 'Participant' is the identification of each person, entered as a random effect.

Models with the best goodness of fit were those which employed shapes or 'area,' see **Table 24**. Models using shapes had the lowest AIC_c scores in three conditions: danger of *family-x*, danger of *family-y*, and beauty of *family-y*. The models using 'area' had the lowest AIC_c score in two conditions: danger of *family-z* and beauty of *family-z*.

The most consistently best fitting model were those using shapes. Shapes as fixed effects also always explained the most variance, R^2_{fixed}. Thus, understanding the shared impressions of novel objects can be uniquely researched by exploring specific shape-parameters. This is not to say that other perceptual variables such as familiarity, spatial frequency, fractality, complexity, size, or curviness do not contribute to the formation of an impression, but rather that there are more ways to understand the relationship between shapes and impressions than previously understood. These results support the notion that shapes map to more than just preferences. This is congruent with prior research, which explored impressions of "innovativeness" amongst car interiors (Leder & Carbon, 2005), or the "sacredness" and "dominance" found amongst certain geometrical figures (Costa & Bonetti, 2016); how shapes map to all impressions of all novel objects could be further elucidated using our approach.

6.1.3.4 Details of Perceptual Variable

6.1.3.4.1 Subjective Familiarity

Preferences amongst abstract stimuli are related to their subjective familiarity (Matlin, 1971; R. B. Zajonc, 1980), and could work by altering the cognitive and perceptual fluency of an object (Reber et al., 2004). Subjective familiarity ratings of each object were collected online using Mturk. 30 participants were recruited, however, 17 participants failed the embedded attention-check and also, provided data without consistency. Thus, the meaningfulness of measuring the subjective familiarity of procedurally generated novel objects should be interpreted with caution. Nevertheless, we proceeded with the analysis by incorporating the remaining data from 13 participants. The procedure was similar to study-2. Each object was presented to participants with the question "How familiar does this object seem to you?" and a scale ranging from 1 (not at all familiar) through 9 (extremely familiar); the object remained visible until the participant responded by keying a number. Each object was selected randomly from a shape-family. The presented order of each shape-family was randomized across participants. After the shape-families were presented once, a break was given before the shape-families were re-evaluated a second time.

Data from the 13 participants who passed the attention check were reliable. We assessed reliability by way of consistency and consensus (the method of leave-one-out correlation with the mean); see study-1's methods for more detail. Significant tests were calculated using Fisher's *r* to *z* transformations and then compared against zero using planned *t*-tests. The level of consistency was found to be significant for each shape-family; *family-x*, $r_M = .62$, $t(12) = 7.71$, $p < .001$, $d = 2.14$; *family-y*, $r_M = .55$, $t(12) = 5.42$, $p < .001$, $d = 1.50$; *family-z*, $r_M = .38$, $t(12) = 3.17$, $p = .008$, $d = .88$. The level of

consensus was also found to be significant for each shape-family—albeit weakly for *family-z*; *family-x*, $r_M = .68$, $t(12) = 7.11$, $p < .001$, $d = 1.97$; *family-y,* $r_M = .46$, $t(12) = 5.03$, $p < .001$, $d = 1.40$; *family-z*, $r_M = .16$, $t(12) = 2.19$, $p = .049$, $d = .61$. Thus, the average subjective familiarity scores for each object reflects the individuals who passed the attention check relatively reliably. These subjective familiarity scores were used to predict impressions.

6.1.3.4.2 Amplitude Spectrum Slope

Fourier spatial frequency amplitude spectrum slopes relate to the sharpness and edges within an image. Images of natural scenes have an average slope of -2.0, while images of objects regularly depict a higher average slope around -2.8 (Redies, Hasenstein, et al., 2007). Natural scenes on average are preferred over images of abstract objects (Vessel & Rubin, 2010). The underlying theoretical basis of these prior observations are separate, the former about vision and the latter about semantics. Nevertheless, the observations are related and therefore we explored the implications of a possible relationship between impression and the amplitude spectrum slope of images of novel object. Specifically, we explored whether the images' amplitude spectrum slope explains how shapes relate to impressions.

The spatial frequency amplitude spectrum slope is calculated by fitting a straight line to a portion of the relationship between the log of the rotationally averaged amplitude spectrum and the log of the spatial frequency (i.e., cycles per image); specifically the portion between 10 and 256 cycles per image—a standard computation procedure for avoiding the effects of image artifacts (Redies, Hasenstein, et al., 2007; Redies, Hänisch,

et al., 2007). The value of the average amplitude spectrum slope for each shape-family in study-2 was congruent with prior research examining objects (Redies, Hasenstein, et al., 2007); *family-x, M* = -2.7; *family-y M* = -2.5; *family-z, M* = -2.8.

6.1.3.4.3 Fractal Dimension

Jackson Pollock's paintings often range in fractal dimensions between 1.1 and 1.9—the same range in which natural scenes are found to also possess (Spehar et al., 2003a). It is possible that each image's fractal dimension meaningfully relates to people's impressions. Therefore, we explored whether the fractal dimension of an image explains how shapes relate to impressions.

We used ImageJ, image processing software, to calculate the fractal dimension of each image (Schneider et al., 2012). To do so, the images were first converted into binary images using the default automatic threshold method. Then each fractal dimension was calculated using the fractal box count analysis with exponential box sizes: 2, 3, 4, 6, 8, 12, 16, 32, 64.

6.1.3.4.4 Image Complexity

In general, people tend to prefer simpler shapes over complex ones (N. Chen et al., 2016), yet the exact relationship between complexity and preferences is the subject of a long debate (e.g., Reber et al., 2004; Sherman et al., 2015). Nevertheless, we explored whether an image's objective complexity explains the relationship between shapes and impressions.

One objective measure of an image's complexity is its compressed file size; a measure that is correlated with subjective measures of complexity (Donderi, 2006). But

unlike subjective complexity ratings, objective complexity ratings are immutable to any objective familiarity (i.e., prior exposure) with a shape (Forsythe et al., 2008). We measured the file size, in bytes, of our images in the jpeg file format.

6.1.3.4.5 Object Volume & Surface Area

Infants as young as 4.5 months old demonstrate the use of an object's size, along with shape, to individuate them (Wilcox, 1999). This basic ability to distinguish size is tied to preferences; people tend to prefer larger shapes, for example (N. Chen et al., 2016; Silvera et al., 2002). People also have in mind a canonical visual size of an object and prefer 2D representations that are congruently sized (Konkle & Oliva, 2011). We explored the possibility that the size of objects relates to peoples' impressions and could therefore explain shapes' effects.

Using the Rhino software, we measured the volume and surface area of each object. Because the objects we test have never occupied physical space, the measurements are unitless. These measurements are proportional to the size of an object as it appears in a 2D image because the rendering process was uniform across all the objects.

6.1.3.4.6 Surface Curvature

People may have a perceptual bias for curvilinear shapes that becomes heightened when primed with a threatening word or fear-inducing video (LoBue, 2014). Smooth curves are also approached more quickly (Palumbo et al., 2015), and are preferred in architectural spaces (Vartanian et al., 2013). We explore whether the surface curvature of each object explains how shapes relate to impressions of novel objects.

Amongst the current literature on curvature's psychological effects, there have been many types of measurements (Gómez-Puerto et al., 2016). Prior research focuses on 2D images and relatively simple contours. Since our stimuli is comprised of 3D geometry, we used a new type of measurement. Specifically, we recorded three variables summarizing the curvature of a 3D surface: Gaussian-max-curvature, Gaussian-min-curvature, and mean-max-curvature. These variables can be recorded using the curvature analysis tool in the Rhino software. The maximum and minimum Gaussian curvature depicts the magnitude of the most extreme synclastic (i.e., bowl-like or dome-like) and anticlastic (i.e., saddle-like) surfaces of an object. Positive values are synclastic, negative values are anticlastic, zero-values represent a flat surface. The maximum mean curvature depicts the most extreme concave/convex surface transition. There are multiple algorithms for estimating the Gaussian and mean curvature (Magid et al., 2007). The algorithm underlying the curvature analysis tool we use in Rhino can be found in chapter 3 of Giovanzana's thesis (2011).

6.2 Chapter 4 Appendix

6.2.1 Study-2 Appendix

Here we present the results of a model based on 12 features. Narrowing down a list of 56 features, we focused on a subset to represent *shapeSpace*. Here we modeled the 12 features using a confirmatory six-dimensional PCA, **Table 25**. The subset of features did well to represent the original *shapeSpace* model. The cumulative variance explained by the six-dimensional model of the 12 features was 98%. Factor congruence across the six dimensions was very high, ranging from .98 to 1.00. Next, we extracted the loadings and performed Procrustes transformation to fit the new loadings to the original *shapeSpace* model, see **Table 26**.

Table 25. Modeling a subset of 12 features using a confirmatory six-dimensional PCA. Then comparing the congruence of its structure to the original model of *shapeSpace* based on 56 features.

(Study-2) Six-dimensional PCA based on 12 features

	PC1	PC2	PC3	PC4	PC5	PC6
dimension	Activity	Fluency	Intuitive Physics	Naturalness	Clearness	Endlessness
average loading	.92	.91	.89	.69	.78	.60
eigenvalue	4.03	2.93	2.55	1.07	.82	.37
variance explained	.34	.24 (.58)	.21 (.79)	.09 (.88)	.07 (.95)	.03 (.98)
factor congruency	1.00	1.00	1.00	1.00	.99	.98

Notes. Correlation based PCA computed using the 'principal' function. Loading is an average of the absolute value of the first three PCs and a direct copy for the last three; it represents the correlation between the scores before PCA and afterwards with the addition of Procrustes transformation to match F56sub12. Eigenvalues = "SS Loadings". Variance explained is the proportion of variance in the data captured by each dimension; the cumulative amount is in parentheses. Factor congruency scores calculated after Procrustes transformation was used to fit F12 to F56sub12.

Table 26. The subset of 12 features are modeled separately and then the loadings are adjusted to fit the original *shapeSpace* based on 56 features.

(Study-2) Subset of 12 features: Loadings

Dimension	Feature	PC1	PC2	PC3	PC4	PC5	PC6
Activity	Plain v. Busy	**.97 (.89)**	.00 (-.11)	.17 (.38)	.04 (-.12)	.09 (.03)	.09 (-.14)
	Rough v. Smooth	**-.87 (-.69)**	-.23 (.12)	-.33 (-.60)	.21 (.32)	-.10 (.12)	-.09 (.10)

	Simple v. Complex	**.94 (.88)**	-.03 (-.16)	.21 (.38)	.19 (.05)	-.01 (.07)	.11 (-.12)
Fluency	Disorganized v. Organized	-.28 (-.53)	**.94 (.59)**	-.11 (.59)	-.02 (-.04)	-.04 (.06)	-.08 (.01)
	Asymmetrical v. Symmetrical	-.19 (-.46)	**.94 (.55)**	-.07 (.62)	-.15 (-.23)	.11 (.05)	-.11 (-.06)
	Negative v. Positive	-.06 (-.23)	**.84 (.59)**	-.19 (.58)	.47 (.30)	-.06 (.40)	.02 (.01)
Intuitive Physics	Weak v. Strong	.02 (-.12)	.20 (-.69)	**.93 (.67)**	-.01 (.02)	-.11 (-.05)	.26 (.20)
	Inelastic v. Elastic	.29 (.46)	-.10 (.70)	**-.92 (-.47)**	.08 (-.07)	.14 (.21)	.13 (.09)
	Light v. Heavy	-.52 (-.55)	.00 (-.72)	**.81 (.28)**	.13 (.10)	.01 (.13)	.19 (.20)
Naturalness	Natural v. Artificial	.44 (.52)	-.16 (-.05)	.02 (.09)	**.69 (.83)**	-.52 (.10)	.01 (-.04)
Clearness	Clear v. Fuzzy	-.08 (.04)	-.38 (-.53)	.27 (-.24)	-.34 (.29)	**.78 (.75)**	.20 (-.06)
Endlessness	Finite v. Endless	.69 (.79)	.02 (.33)	-.39 (.06)	-.08 (-.13)	-.06 (.03)	**.60 (.48)**

Notes. Loadings are correlations between features and dimensions. Procrustes transformation was performed to fit the loadings to shapeSpace (based on 56 features). In parentheses is the untransformed factor loadings.

6.2.2 Study-3 Appendix

6.2.2.1 Consistency & Consensus

We report a series of planned two-tailed *t*-tests against zero for each feature's consistency and consensus. We report this for every shape-family. For BB1 see **Table 34**; for BB2 see **Table 35**; for EG1 see **Table 36**; for EG2 see **Table 37**; for SA1 see **Table 38**; for SA2 see **Table 39**.

Table 27. Shape-family BB1 feature reliability

(Study-3) BB1: Consistency & Consensus

Feature	Measure	*df*	*T*	*p*	*d*	r_M
Asymmetrical v. Symmetrical	Consistency	18	4.35	< .001	1.00	.33
	Consensus-U	18	8.71	< .001	2.00	.54
	Consensus-F	13	9.15	< .001	2.45	.59
Clear v. Fuzzy	Consistency	18	4.32	< .001	.99	.37
	Consensus-U	18	3.97	< .001	.91	.34
	Consensus-F	12	3.96	.002	1.10	.43
Disorganized v. Organized	Consistency	18	4.81	< .001	1.10	.36
	Consensus-U	18	3.53	.002	.81	.24
	Consensus-F	13	4.32	< .001	1.15	.32
Finite v. Endless	Consistency	18	5.01	< .001	1.15	.39
	Consensus-U	18	4.87	< .001	1.12	.34
	Consensus-F	14	4.32	< .001	1.12	.33

Inelastic v. Elastic*	Consistency	18	4.19	< .001	.96	.43
	Consensus-U	18	5.39	< .001	1.24	.52
	Consensus-F	15	5.70	< .001	1.42	.59
Light v. Heavy	Consistency	18	7.74	< .001	1.78	.53
	Consensus-U	18	8.47	< .001	1.93	.57
	Consensus-F	16	7.45	< .001	1.81	.55
Natural v. Artificial	Consistency	18	5.49	< .001	1.26	.39
	Consensus-U	18	4.18	< .001	.96	.40
	Consensus-F	13	3.82	.002	1.02	.49
Negative v. Positive	Consistency	18	6.57	< .001	1.51	.47
	Consensus-U	18	5.90	< .001	1.35	.49
	Consensus-F	15	6.42	< .001	1.60	.53
Plain v. Busy	Consistency	18	8.64	< .001	1.98	.52
	Consensus-U	18	10.81	< .001	2.48	.70
	Consensus-F	17	12.88	< .001	3.04	.72
Rough v. Smooth*	Consistency	18	5.72	< .001	1.31	.56
	Consensus-U	18	8.52	< .001	1.95	.72
	Consensus-F	16	11.67	< .001	2.83	.76
Simple v. Complex	Consistency	18	9.37	< .001	2.15	.58
	Consensus-U	18	13.03	< .001	2.99	.69
	Consensus-F	18	13.03	< .001	2.99	.69
Weak v. Strong	Consistency	18	9.46	< .001	2.17	.59
	Consensus-U	18	8.01	< .001	1.84	.68
	Consensus-F	17	7.97	< .001	1.88	.69

*Notes. BB1. Planned two-tailed t-tests against zero; d = Cohen's effect size; r_M = average Pearson's r correlation. *feature is flipped relative to study-2 when tested in study-3.*

Table 28. Shape-family BB2 feature reliability.

(Study-3) BB2: Consistency & Consensus

Feature	Measure	*df*	*t*	*p*	*d*	r_M
Asymmetrical v. Symmetrical	Consistency	18	5.82	< .001	1.34	.40
	Consensus-U	18	9.45	< .001	2.17	.54
	Consensus-F	14	8.05	< .001	2.08	.55
Clear v. Fuzzy	Consistency	18	6.19	< .001	1.42	.53
	Consensus-U	18	2.61	.018	.60	.41
	Consensus-F	16	3.16	.006	.77	.49
Disorganized v. Organized	Consistency	18	6.73	< .001	1.54	.45
	Consensus-U	18	3.46	.003	.79	.32
	Consensus-F	16	3.15	.006	.76	.34
Finite v. Endless	Consistency	18	6.96	< .001	1.60	.52
	Consensus-U	18	6.90	< .001	1.58	.56
	Consensus-F	16	6.22	< .001	1.51	.56
Inelastic v. Elastic*	Consistency	18	4.76	< .001	1.09	.37
	Consensus-U	18	5.61	< .001	1.29	.49
	Consensus-F	15	5.32	< .001	1.34	.53
Light v. Heavy	Consistency	18	9.18	< .001	2.11	.63
	Consensus-U	18	10.16	< .001	2.33	.74
	Consensus-F	17	9.94	< .001	2.34	.75
Natural v. Artificial	Consistency	18	4.70	< .001	1.08	.41
	Consensus-U	18	8.73	< .001	2.00	.53
	Consensus-F	16	7.89	< .001	1.91	.53
Negative v. Positive	Consistency	18	5.98	< .001	1.37	.50
	Consensus-U	18	1.52	.147	.35	.17
	Consensus-F	16	1.44	.169	.35	.18

Plain v. Busy	Consistency	18	12.16	< .001	2.79	.67
	Consensus-U	18	12.01	< .001	2.75	.72
	Consensus-F	18	12.01	< .001	2.75	.72
Rough v. Smooth*	Consistency	18	7.76	< .001	1.78	.49
	Consensus-U	18	7.45	< .001	1.71	.59
	Consensus-F	16	6.66	< .001	1.62	.60
Simple v. Complex	Consistency	18	12.15	< .001	2.79	.63
	Consensus-U	18	9.70	< .001	2.22	.73
	Consensus-F	18	9.70	< .001	2.22	.73
Weak v. Strong	Consistency	18	7.80	< .001	1.79	.58
	Consensus-U	18	9.17	< .001	2.10	.68
	Consensus-F	17	9.50	< .001	2.24	.70

*Notes. BB2. Planned two-tailed t-tests against zero; d = Cohen's effect size; r_M = average Pearson's r correlation. *feature is flipped relative to study-2 when tested in study-3.*

Table 29. Shape-family EG1 feature reliability.

(Study-3) EG1: Consistency & Consensus

Feature	Measure	*df*	*t*	*p*	*d*	r_M
Asymmetrical v. Symmetrical	Consistency	17	6.01	< .001	1.42	.61
	Consensus-U	17	14.68	< .001	3.46	.81
	Consensus-F	16	17.79	< .001	4.32	.82
Clear v. Fuzzy	Consistency	17	3.58	.002	.84	.31
	Consensus-U	17	5.89	< .001	1.39	.48
	Consensus-F	13	7.95	< .001	2.12	.56
Disorganized v. Organized	Consistency	17	7.64	< .001	1.80	.62
	Consensus-U	17	10.33	< .001	2.43	.75
	Consensus-F	16	15.05	< .001	3.65	.78
Finite v. Endless	Consistency	17	4.66	< .001	1.10	.43
	Consensus-U	17	5.35	< .001	1.26	.42
	Consensus-F	12	5.43	< .001	1.51	.48
Inelastic v. Elastic*	Consistency	17	2.61	.018	.62	.33
	Consensus-U	17	2.98	.008	.70	.34
	Consensus-F	12	2.32	.039	.64	.35
Light v. Heavy	Consistency	17	6.96	< .001	1.64	.51
	Consensus-U	17	5.16	< .001	1.22	.40
	Consensus-F	16	5.29	< .001	1.28	.41
Natural v. Artificial	Consistency	17	3.82	.001	.90	.30
	Consensus-U	17	.15	.880	.04	.01
	Consensus-F	10	-.99	.344	.30	-.09
Negative v. Positive	Consistency	17	2.87	.011	.68	.33
	Consensus-U	17	3.90	.001	.92	.41
	Consensus-F	13	4.55	< .001	1.22	.53
Plain v. Busy	Consistency	17	9.31	< .001	2.19	.54
	Consensus-U	17	19.37	< .001	4.57	.71
	Consensus-F	17	19.37	< .001	4.57	.71
Rough v. Smooth*	Consistency	17	11.86	< .001	2.80	.59
	Consensus-U	17	8.46	< .001	1.99	.65
	Consensus-F	17	8.46	< .001	1.99	.65
Simple v. Complex	Consistency	17	11.18	< .001	2.64	.58
	Consensus-U	17	13.67	< .001	3.22	.74
	Consensus-F	17	13.67	< .001	3.22	.74
Weak v. Strong	Consistency	17	5.46	< .001	1.29	.47
	Consensus-U	17	4.50	< .001	1.06	.32
	Consensus-F	15	4.71	< .001	1.18	.39

*Notes. EG1. Planned two-tailed t-tests against zero; d = Cohen's effect size; r_M = average Pearson's r correlation. *feature is flipped relative to study-2 when tested in study-3.*

Table 30. Shape-family EG2 feature reliability.

(Study-3) EG2: Consistency & Consensus

Feature	Measure	*df*	*t*	*p*	*d*	r_M
Asymmetrical v. Symmetrical	Consistency	16	5.14	< .001	1.25	.43
	Consensus-U	16	4.32	< .001	1.06	.41
	Consensus-F	15	4.21	< .001	1.05	.43
Clear v. Fuzzy	Consistency	16	7.17	< .001	1.74	.64
	Consensus-U	16	7.72	< .001	1.87	.76
	Consensus-F	15	11.25	< .001	2.81	.80
Disorganized v. Organized	Consistency	16	6.50	< .001	1.58	.54
	Consensus-U	16	5.71	< .001	1.39	.63
	Consensus-F	15	5.55	< .001	1.39	.64
Finite v. Endless	Consistency	16	3.24	.005	.79	.33
	Consensus-U	16	4.07	< .001	.99	.44
	Consensus-F	13	3.84	.002	1.03	.44
Inelastic v. Elastic*	Consistency	16	3.16	.006	.77	.44
	Consensus-U	16	1.56	.139	.38	.24
	Consensus-F	12	.89	.390	.25	.18
Light v. Heavy	Consistency	16	9.59	< .001	2.33	.69
	Consensus-U	16	9.38	< .001	2.28	.81
	Consensus-F	16	9.38	< .001	2.28	.81
Natural v. Artificial	Consistency	16	7.75	< .001	1.88	.59
	Consensus-U	16	3.72	.002	.90	.57
	Consensus-F	15	3.83	.002	.96	.60
Negative v. Positive	Consistency	16	4.43	< .001	1.07	.42
	Consensus-U	16	4.27	< .001	1.04	.43
	Consensus-F	13	4.31	< .001	1.15	.51
Plain v. Busy	Consistency	16	6.93	< .001	1.68	.59
	Consensus-U	16	6.18	< .001	1.50	.60
	Consensus-F	14	5.27	< .001	1.36	.60
Rough v. Smooth*	Consistency	16	11.38	< .001	2.76	.72
	Consensus-U	16	13.71	< .001	3.33	.88
	Consensus-F	16	13.71	< .001	3.33	.88
Simple v. Complex	Consistency	16	10.00	< .001	2.43	.56
	Consensus-U	16	5.72	< .001	1.39	.52
	Consensus-F	16	5.72	< .001	1.39	.52
Weak v. Strong	Consistency	16	4.74	< .001	1.15	.47
	Consensus-U	16	1.78	.094	.43	.30
	Consensus-F	14	1.60	.132	.41	.30

*Notes. EG2. Planned two-tailed t-tests against zero; d = Cohen's effect size; r_M = average Pearson's r correlation. *feature is flipped relative to study-2 when tested in study-3.*

Table 31. Shape-family SA1 feature reliability.

(Study-3) SA1: Consistency & Consensus

Feature	Measure	*df*	*t*	*p*	*d*	r_M
Asymmetrical v. Symmetrical	Consistency	17	4.27	< .001	1.01	.39
	Consensus-U	17	5.15	< .001	1.21	.48
	Consensus-F	12	7.01	< .001	1.95	.58

Clear v. Fuzzy	Consistency	17	3.03	.008	.71	.38
	Consensus-U	17	3.33	.004	.79	.38
	Consensus-F	15	3.20	.006	.80	.40
Disorganized v. Organized	Consistency	17	7.20	< .001	1.70	.59
	Consensus-U	17	8.57	< .001	2.02	.68
	Consensus-F	15	8.96	< .001	2.24	.72
Finite v. Endless	Consistency	17	3.80	.001	.90	.39
	Consensus-U	17	1.40	.179	.33	.14
	Consensus-F	14	1.10	.290	.28	.14
Inelastic v. Elastic*	Consistency	17	2.17	.045	.51	.38
	Consensus-U	17	5.35	< .001	1.26	.67
	Consensus-F	11	11.58	< .001	3.34	.84
Light v. Heavy	Consistency	17	4.30	< .001	1.01	.48
	Consensus-U	17	4.19	< .001	.99	.42
	Consensus-F	12	4.77	< .001	1.32	.57
Natural v. Artificial	Consistency	17	6.92	< .001	1.63	.55
	Consensus-U	17	9.69	< .001	2.28	.70
	Consensus-F	15	11.99	< .001	3.00	.74
Negative v. Positive	Consistency	17	4.37	< .001	1.03	.48
	Consensus-U	17	2.77	.013	.65	.39
	Consensus-F	12	3.16	.008	.88	.55
Plain v. Busy	Consistency	17	7.01	< .001	1.65	.59
	Consensus-U	17	10.14	< .001	2.39	.72
	Consensus-F	16	11.33	< .001	2.75	.74
Rough v. Smooth*	Consistency	17	11.26	< .001	2.65	.76
	Consensus-U	17	8.63	< .001	2.03	.88
	Consensus-F	17	8.63	< .001	2.03	.88
Simple v. Complex	Consistency	17	7.87	< .001	1.86	.60
	Consensus-U	17	3.41	.003	.80	.23
	Consensus-F	16	11.42	< .001	2.77	.74
Weak v. Strong	Consistency	17	4.58	< .001	1.08	.48
	Consensus-U	17	3.41	.003	.80	.23
	Consensus-F	14	2.73	.016	.71	.26

*Notes. SA1. Planned two-tailed t-tests against zero; d = Cohen's effect size; r_M = average Pearson's r correlation. *feature is flipped relative to study-2 when tested in study-3.*

Table 32. Shape-family SA2 feature reliability.

(Study-3) SA2: Consistency & Consensus

Feature	Measure	*df*	*t*	*p*	*d*	r_M
Asymmetrical v. Symmetrical	Consistency	15	3.05	.008	.76	.29
	Consensus-U	15	8.38	< .001	2.10	.53
	Consensus-F	11	6.98	< .001	2.02	.55
Clear v. Fuzzy	Consistency	15	4.15	< .001	1.04	.49
	Consensus-U	15	3.94	.001	.98	.62
	Consensus-F	12	4.28	.001	1.19	.71
Disorganized v. Organized	Consistency	15	3.28	.005	.82	.39
	Consensus-U	15	6.83	< .001	1.71	.55
	Consensus-F	12	8.21	< .001	2.28	.62
Finite v. Endless	Consistency	15	4.70	< .001	1.18	.43
	Consensus-U	15	3.14	.007	.79	.53
	Consensus-F	11	2.81	.017	.81	.58
Inelastic v. Elastic*	Consistency	15	2.81	.013	.70	.31
	Consensus-U	15	2.24	.040	.56	.32
	Consensus-F	7	1.77	.120	.63	.46

Light v. Heavy	Consistency	15	7.67	< .001	1.92	.49
	Consensus-U	15	8.40	< .001	2.10	.56
	Consensus-F	15	8.40	< .001	2.10	.56
Natural v. Artificial	Consistency	15	4.27	< .001	1.07	.40
	Consensus-U	15	.70	.492	.18	.10
	Consensus-F	12	.90	.387	.25	.15
Negative v. Positive	Consistency	15	1.68	.114	.42	.16
	Consensus-U	15	3.74	.002	.94	.36
	Consensus-F	7	4.09	.005	1.45	.52
Plain v. Busy	Consistency	15	11.42	< .001	2.85	.66
	Consensus-U	15	21.54	< .001	5.38	.87
	Consensus-F	15	21.54	< .001	5.38	.87
Rough v. Smooth*	Consistency	15	12.58	< .001	3.15	.65
	Consensus-U	15	22.86	< .001	5.72	.84
	Consensus-F	15	22.86	< .001	5.72	.84
Simple v. Complex	Consistency	15	10.06	< .001	2.52	.64
	Consensus-U	15	16.96	< .001	4.24	.85
	Consensus-F	15	16.96	< .001	4.24	.85
Weak v. Strong	Consistency	15	3.28	.005	.82	.34
	Consensus-U	15	4.04	.001	1.01	.36
	Consensus-F	11	4.31	.001	1.25	.46

*Notes. SA2. Planned two-tailed t-tests against zero; d = Cohen's effect size; r_M = average Pearson's r correlation. *feature is flipped relative to study-2 when tested in study-3.*

6.2.2.2 Loadings

We calculated the correlation between features and dimensions (i.e., loadings). Procrustes transformation was used to maximize the fit, while preserving the relational structure across the loadings. The transformations were used to fit the confirmatory models to the original exploratory model of *shapeSpace*. We first report the loadings of a model representing impressions across all shape-families, see **Table 33**. Then for each shape-family; for BB1 see **Table 34**; for BB2 see **Table 35**; for EG1 see **Table 36**; for EG2 see **Table 37**; for SA1 see **Table 38**; for SA2 see **Table 39**.

Table 33. The loadings of a six-dimensional confirmatory PCA that projects impressions of different shape-families together in the same space.

(Study-3) Across shape-families: Loadings

Dimension	Feature	PC1	PC2	PC3	PC4	PC5	PC6
Activity	Plain v. Busy	**.85 (.88)**	-.04 (.16)	.46 (.16)	-.07 (-.27)	.11 (.19)	.15 (-.21)
	Rough v. Smooth*	**-.72 (-.87)**	.25(.12)	-.41 (.13)	-.09 (.26)	.37 (-.17)	-.14 (-.19)
	Simple v. Complex	**.80 (.85)**	-.01 (.22)	.49 (.22)	.00 (-.24)	.26 (.06)	.10 (-.28)
Fluency	Disorganized v. Organized	-.26 (-.50)	**.84 (.71)**	-.21 (-.33)	.25 (-.12)	-.14 (-.06)	-.03 (.18)
	Asymmetrical v. Symmetrical	-.06 (-.31)	**.92 (.83)**	-.08 (-.27)	.17 (-.18)	-.04 (.01)	.00 (.04)
	Negative v. Positive	-.36 (-.61)	**.74 (.62)**	-.21 (.05)	-.22 (.12)	.34 (.15)	.05 (-.29)
Intuitive Physics	Weak v. Strong	.05 (.64)	.15 (.35)	**.82 (-.31)**	.08 (.48)	-.19 (.10)	.37 (.07)
	Inelastic v. Elastic*	-.29 (-.62)	-.21 (-.30)	**-.75 (.47)**	-.34 (.06)	.03 (.44)	.38 (.26)
	Light v. Heavy	.07 (.67)	-.06 (.22)	**.77 (.01)**	-.07 (.60)	.03 (.21)	.55 (.00)
Naturalness	Natural v. Artificial	.42 (.76)	-.15 (-.05)	.55 (-.45)	**.32 (-.11)**	-.51 (-.11)	-.04 (.26)
Clearness	Clear v. Fuzzy	.15 (.36)	-.19 (.26)	.16 (.69)	.57 (.20)	**.66 (-.51)**	.35 (.10)
Endlessness	Finite v. Endless	.46 (.23)	.38 (.67)	-.14 (.57)	.29 (-.17)	.33 (.13)	**.62 (.27)**

Notes. Loadings are correlations between features and dimensions. Procrustes transformation was performed to adjust the dimensions to best fit study-2's shapeSpace. Untransformed loadings are displayed in parentheses.

Table 34. A confirmatory model of shape-family BB1's loadings in relation to *shapeSpace.*

(Study-3) BB1: Loadings

Dimension	Feature	PC1	PC2	PC3	PC4	PC5	PC6
Activity	Plain v. Busy	.81	.21	.46	.14	.02	.23
	Rough v. Smooth*	-.69	.42	-.45	-.24	-.06	-.24
	Simple v. Complex	.79	.17	.45	.15	.04	.30
Fluency	Disorganized v. Organized	.01	.94	-.21	.17	-.15	.03
	Asymmetrical v. Symmetrical	-.18	.89	-.25	-.11	.20	.20
	Negative v. Positive	-.09	.77	-.53	-.17	-.02	-.26
Intuitive Physics	Weak v. Strong	-.03	.00	.98	-.03	-.08	.11
	Inelastic v. Elastic*	.03	-.01	-.94	.17	-.08	.26
	Light v. Heavy	.66	-.14	.48	.16	.00	.50
Naturalness	Natural v. Artificial	.41	-.59	.45	.43	.11	.15
Clearness	Clear v. Fuzzy	.50	-.44	.52	.34	.27	.19
Endlessness	Finite v. Endless	.71	.42	-.19	-.14	.20	.45

Notes. Loadings are correlations between features and dimensions. Procrustes transformation was performed to adjust the dimensions to best fit study-2's shapeSpace.

Table 35. A confirmatory model of shape-family BB2's loadings in relation to *shapeSpace.*

(Study-3) BB2: Loadings

Dimension	Feature	PC1	PC2	PC3	PC4	PC5	PC6
Activity	Plain v. Busy	.76	.28	.45	.19	.02	.29
	Rough v. Smooth*	-.72	.09	-.68	-.04	.01	-.03
	Simple v. Complex	.71	.33	.47	.22	.01	.33
Fluency	Disorganized v. Organized	-.07	.89	-.29	-.15	-.28	.07
	Asymmetrical v. Symmetrical	.22	.95	.13	-.07	.10	.07
	Negative v. Positive	.24	.79	-.49	.26	.01	.07
Intuitive Physics	Weak v. Strong	.42	.11	.81	.18	-.04	.34
	Inelastic v. Elastic*	-.34	.06	-.91	-.05	-.08	-.06
	Light v. Heavy	.31	-.03	.84	.23	-.06	.37
Naturalness	Natural v. Artificial	.42	-.25	.81	.25	-.12	.07
Clearness	Clear v. Fuzzy	.26	-.45	.77	.18	.25	.15
Endlessness	Finite v. Endless	.56	.66	.15	-.01	-.03	.45

Notes. Loadings are correlations between features and dimensions. Procrustes transformation was performed to adjust the dimensions to best fit study-2's shapeSpace.

Table 36. A confirmatory model of shape-family EG1's loadings in relation to *shapeSpace.*

(Study-3) EG1: Loadings

Dimension	Feature	PC1	PC2	PC3	PC4	PC5	PC6

Activity	Plain v. Busy	.74	-.46	-.02	.11	.41	.20
	Rough v. Smooth*	-.82	-.12	-.27	.16	-.31	-.32
	Simple v. Complex	.64	-.64	-.09	.23	.30	.10
Fluency	Disorganized v. Organized	-.38	.87	.15	-.06	-.24	.01
	Asymmetrical v. Symmetrical	-.42	.85	.10	.04	-.24	-.07
	Negative v. Positive	-.34	.87	.20	.17	-.13	-.13
Intuitive Physics	Weak v. Strong	-.35	.68	.56	-.08	-.27	.04
	Inelastic v. Elastic*	.61	-.66	-.17	.00	.24	.22
	Light v. Heavy	-.11	-.13	.77	.43	.15	.41
Naturalness	Natural v. Artificial	-.15	.20	.26	.57	-.73	-.12
Clearness	Clear v. Fuzzy	.80	-.26	.16	-.17	.31	.32
Endlessness	Finite v. Endless	.67	.20	-.06	-.28	.26	.59

Notes. Loadings are correlations between features and dimensions. Procrustes transformation was performed to adjust the dimensions to best fit study-2's shapeSpace.

Table 37. A confirmatory model of shape-family EG2's loadings in relation to *shapeSpace.*

(Study-3) EG2: Loadings

Dimension	Feature	PC1	PC2	PC3	PC4	PC5	PC6
Activity	Plain v. Busy	.71	-.32	.60	.06	-.15	.00
	Rough v. Smooth*	-.62	.58	-.45	.09	.20	.10
	Simple v. Complex	.47	.00	.75	.31	.01	.34
Fluency	Disorganized v. Organized	-.51	.67	-.41	.25	.20	.08
	Asymmetrical v. Symmetrical	-.35	.47	-.69	-.26	.29	.01
	Negative v. Positive	-.47	.57	-.33	.27	.48	.13
Intuitive Physics	Weak v. Strong	.41	-.39	.75	-.11	-.29	-.01
	Inelastic v. Elastic*	-.08	.10	-.93	-.28	-.09	-.15
	Light v. Heavy	.14	-.08	.87	.32	-.13	.29
Naturalness	Natural v. Artificial	.48	-.56	.57	.04	-.34	-.06
Clearness	Clear v. Fuzzy	-.26	.66	-.28	.28	.51	.23
Endlessness	Finite v. Endless	-.22	.56	-.48	.17	.44	.42

Notes. Loadings are correlations between features and dimensions. Procrustes transformation was performed to adjust the dimensions to best fit study-2's shapeSpace.

Table 38. A confirmatory model of shape-family SA1's loadings in relation to *shapeSpace.*

(Study-3) SA1: Loadings

Dimension	Feature	PC1	PC2	PC3	PC4	PC5	PC6
Activity	Plain v. Busy	.87	-.36	.26	-.13	-.14	-.06
	Rough v. Smooth*	-.82	.04	-.20	.03	.41	.31
	Simple v. Complex	.88	-.29	.32	-.07	-.12	-.07
Fluency	Disorganized v. Organized	-.16	.86	-.18	.05	-.38	-.20

	Asymmetrical v. Symmetrical	-.18	.97	.00	-.01	.01	-.15
	Negative v. Positive	-.71	.26	-.17	-.09	.26	.54
Intuitive Physics	Weak v. Strong	-.64	.45	.46	.03	.16	.36
	Inelastic v. Elastic*	-.83	.18	-.22	-.03	.29	.35
	Light v. Heavy	-.67	.15	.39	.20	.30	.47
Naturalness	Natural v. Artificial	.73	.06	.23	.18	-.45	-.40
Clearness	Clear v. Fuzzy	-.71	-.29	.10	.33	.50	.20
Endlessness	Finite v. Endless	-.57	.54	-.35	.26	.19	.39

Notes. Loadings are correlations between features and dimensions. Procrustes transformation was performed to adjust the dimensions to best fit study-2's shapeSpace.

Table 39. A confirmatory model of shape-family SA2's loadings in relation to *shapeSpace.*

(Study-3) SA2: Loadings

Dimension	Feature	PC1	PC2	PC3	PC4	PC5	PC6
Activity	Plain v. Busy	.70	-.19	.61	.05	.12	.26
	Rough v. Smooth*	-.76	.19	-.55	-.07	-.04	-.24
	Simple v. Complex	.71	-.26	.57	.11	.03	.28
Fluency	Disorganized v. Organized	-.57	.62	-.48	-.12	-.16	-.10
	Asymmetrical v. Symmetrical	.17	.89	.07	-.14	.29	-.26
	Negative v. Positive	-.67	.50	-.51	.04	-.06	-.09
Intuitive Physics	Weak v. Strong	.51	.05	.80	-.02	-.17	.23
	Inelastic v. Elastic*	-.75	.18	-.55	.01	-.17	-.23
	Light v. Heavy	-.11	-.36	.81	.29	.11	.32
Naturalness	Natural v. Artificial	.72	-.36	.52	.27	-.02	.07
Clearness	Clear v. Fuzzy	.77	-.21	.51	.09	.08	.25
Endlessness	Finite v. Endless	.74	-.23	.49	.01	.18	.34

Notes. Loadings are correlations between features and dimensions. Procrustes transformation was performed to adjust the dimensions to best fit study-2's shapeSpace.